A Guide that Opens Science

Gjalt-Jorn Peters Tuğçe Varol

2025-03-16

Table of contents

Preface

Welcome to this **Guide** that **Opens Science** (by Gjalt-Jorn Peters & Tuğçe Varol). This is meant as a brief introduction and reference for people exploring Open Science principles and practices, written from the perspective of health psychology research.

If you would like to share this guide with somebody, you can let them scan the QR-code shown below, or point them to guide.opens.science.

You can cite this as:

Peters, G.-J., & Varol, T. (2024). A Guide that Opens Science. Zenodo. https://doi.org/mzbh

1 Citizen Science

Citizen science is scientific research conceptualized, designed, planned, conducted, analysed, and/or published by people who are not employed as scientists (for example, at an academic institution). Common forms of citizen science are participatory action research, including societal stakeholders in a sounding board, and crowdsourcing data collection. Citizen science can help to increase the public's understanding of and appreciation for science.

FORRT has extensive resources about citizen science at https://forrt.org/glossary/citizen-science/.

2 Open Infra

A vital part of open science, but one that is paradoxically easy to overlook, is open infrastructure.

It is easy to overlook because infrastructure (or 'infra' for short) is so ubiquitous — the term refers to the tools you use for "doing science". This includes hardware (such as eye-tracking devices or EEG equipment), software (such as applications to administer online studies or to analyse data), but also conceptual tools (such as tools to develop theory or document and justify decisions).

Open infrastructure is infrastructure that is open source (or more accurately, "free"), which means you have the following freedoms:

- The freedom to run the program as you wish, for any purpose (freedom 0).
- The freedom to study how the program works, and change it so it does your computing as you wish (freedom 1). Access to the source code is a precondition for this.
- The freedom to redistribute copies so you can help your neighbor (freedom 2).
- The freedom to distribute copies of your modified versions to others (freedom 3). By doing this you can give the whole community a chance to benefit from your changes. Access to the source code is a precondition for this.

In addition, for Open Science, it's important the infrastructure is community-owned and community-governed, instead of it being governed by a single company, government, or institution.

2.1 How to use Open Infra

Using Open Infra is simultaneously easy and hard. It is easy because it usually only involves switching software (see the next section for examples). It is hard because on the one hand, learning new software can take time and energy; and on the other hand, you may be ahead of your institution's policies, so officially, maybe only closed science software may be supported. Depending on your situation and on the specific software, you may be able to make the switch (see the next section) or you may have to advocate at your institution (see the section after that).

2.2 Open Infra Substitutions

This list is based on research in health psychology, so it will probably not be applicable to all researchers.

Task or Function	Software type	Closed Science	Open Science
Managing citations or bibliographies	Reference Manager	EndNote, Mendeley	Zotero, JabRef
Writing documents	Text Processor	Google Docs, MS Word, Apple Pages	LibreOffice Writer, Typst, Markdown, Quarto
Collaboration and version control	Git Forge	GitHub	Codeberg, GitLab
Online Studies	Data collection	Qualtrics, SurveyMonkey	LimeSurvey, {formr}
Statistical analyses	Data Analysis	SPSS, SAS, Stata	jamovi, JASP, R & RStudio, Python
Qualitative analyses	Data Analysis	NVivo, Atlas.ti, Dedoose, MaxQDA	ROCK, QualCoder, Taguette, ReQual
Mock-ups and wireframes	Design	Figma	Penpot
Scientific figures	Design	Adobe Illustrator, Canva	InkScape

2.3 Institutions' responsibilities

However, your institution may not yet have updated their infrastructure to open infrastructure. Implementing new software requires planning and training of staff, as well as potentially overcoming resistance among staff, who can be reluctant to learn new applications (e.g. due to work pressure, a lack of perceived support, various degrees of IT anxiety, aversion to change, etc).

As a consequence, institions can choose to give open infrastructure a low priority. This is ironic, as the longer closed infrastructure is used, the more expensive it will be to transition to open infrastructure later. For example, interoperability, one of the four characteristics of FAIR open data, usually requires open standards, and these are generally implemented in open infrastructure but often missing from closed infrastructure (since propietary software is often developed by corporations with an interest in locking customers into their ecosystems, something made considerably harder by open standards).

As researchers, you can point out the responsibilities of your institutions through a number of channels:

- Faculties often have some council or committee dedicated to research policy. You can contact your representatives on this council to ask them to bring this forward.
- Similarly, faculties generally have democratic bodies such as faculty councils. You can contact your representatives here as well.
- You can also issue a request to your university's IT department.
- You can also point to the UNESCO recommendation on Open Science (UNESCO 2021).

3 Preregistration

Preregistering is a form of registration: freezing a snapshot of information about a study, usually accompanied by the files at that stage of the process. Specifically, a preregistration is a registration done before ('pre') the relevant step in the process: typically, before data collection.

Preregistration forms have been argued to have a number of epistemological benefits, but regardless of those advantages, they are also a great help in planning your study. They are a great way to get your team on the same page regarding a number of relevant aspects of your study.

There's a by now classic blog post introducing (pre)registrations at https://www.cos.io/blog/preregistration-plan-not-prison. In addition, FORRT lists a number of excellent resources: https://forrt.org/forrt_clusters/preregistration/.

There exist many preregistration forms. Over time, you will discover which ones are the most useful for the types of studies you do. If you're just starting out preregistering, we recommend using the Open Science Framework. That links the preregistration to your project's repository (and so, to the preprint once you posted that).

The OSF has a number of forms built-in, but many forms have also been developed that are not yet integrated in OSF. If you want to use one of those forms, you can complete them on your PC, export to PDF, and then attach to the Open-Ended registration form. This is just a text box, optionally accompanied by one or more files.

As a starting point, the following forms can be considered.

- If you find no other form that fits, then as a **fall-back form** the Inclusive General-Purpose Registration Form can be used. This form was designed to be inclusive to all designs and disciplines. That also means it has fewer benefits to your planning than a more specialized form.
- For quantitative studies there are two general-purpose forms. It's best to use an even more specialized one (so do read on), but if you can't find one, these are good fall-back forms if they apply. These are the **OSF preregistration** and the **Psychological Research Preregistration-Quantitative** (aka PRP-QUANT) Template. Both are on OSF, too.
- For Experience Sampling Method studies, there is the **Template for Registration of Studies Using Experience-Sampling Methods**. This is not yet integrated in OSF or {preregr}, so to use it, you attach it to the OSF Open-Ended Registration Form.

- If you do a qualitative study, there are two forms, too. There is the less specific **Qualitative Preregistration Template**, and the **Preregistration Template for Qualitative and Quantitative Ethnographic Studies**. The latter one is more extensive, and so helps you consider and discuss more things before you start data collection.
- If you do a systematic review, such as a scoping review, a meta-analysis, or another type of systematic review, you can use the **Generalized Systematic Review Registration Form**.
- If you do a study with secondary data, you can use the **Preregistration Template for Secondary Data Analysis**.

Whichever form you use, you can extend it with add-ons. So far, one exists.

- The **Inclusivity & Diversity Add-on for preregistration forms** helps you address another Open Science area: inclusivity and diversity, specifically relating mostly to sampling and data registration.

4 Open Data

Publishing open data comes in many shapes: in other words, in the buffet of open data sharing possibilities, there's bound to be something you like. In general, like with other open science practices, it's good to keep the "as open as possible, as closed as necessary" principle in mind.

4.1 Privacy

From this, it follows that there are circumstances where you cannot publish data at all. This is in cases where anonymization is impossible. The General Data Protection Regulation (GDPR) defines personal data as data about an identified or identifiable person. Anonymous data, then, are data that are not about an identified or identifiable person.

A person is identified when it is clear who the person is. A person is identifiable when it is possible, with reasonable effort, to find out who they are. For example, if I tell you that a person's age is 62, that they identify as Buddhist, and that they live in a rural location, that doesn't tell you anything. When I also add that they identify as nonbinary, lost a leg in a war, and if I name the village where they live with a population of 500 inhabitants, they become identifiable.

For research, this means that whether data are personal data are the consequence of three things:

1) how much data you collected;
2) in which 'resolution' you collected the data; and
3) your sampling frame.

The more data you collect, the easier identification becomes. For example, video data is so rich ('much') that it's often considered personal data. Answers to questions in an online questionnaire, however, lend themselves much to identification.

If you register age in years, the resolution of your data is higher than if you use decade-wide categories. In the latter situation, age is unlikely to contribute to identifiability of your participants.

Finally, if you sample from all residents in a country with millions of residents, the risk of identification is much lower than when you sample from, for example, all first-year psychology students.

In addition, there's also the k-anonimity approach to anonymity: that's explained in a bit more detail here.

Ideally, then, you sample as widely as possible, and register as little as possible. In cases where you're not certain, make sure you consult your organization's privacy officer.

4.2 Being FAIR

As anybody who's ever tried to work with published data can attest, publishing data isn't just a matter of uploading some files. The files have to be intelligible to others (and maybe more importantly, to future you), with as little effort as possible.

To cover your bases in this respect, the FAIR acronym is helpful. It captures the mandate to publish data that are Findable, Accessible, Interoperable, and Reusable.

4.2.1 Reusable

Of these, the R (reuse) is easiest to realize, so we'll start there. Because you data cannot own data you collect this means that most cases everybody can reuse your data. But, it helps to be explicit about this by attaching an explicit license. The Open Database License (ODbL) is a good choice if you don't want to dive into this.

4.2.2 Interoperable

The I (interoperable) is hardest for psychology. That is, it has a very minimal interpretation which just means that the data has to be stored in files in open formats (see Accessibility); but true interoperability requires that the data set can be understood by machines. This requires unique identifiers for columns that can be widely understood by machines (e.g. linked back to constructs). Therefore, in practice, you can ignore this for now — which makes it easy, too If you would like to work on this, see the link to the blog post in the Findable section.

4.2.3 Accessible

To be accessible, the data has to be stored in a file format that is, ideally, an open format (also see Chapter 2).

4.2.4 Findable

To make data findable, it has to be deposited in a well-known repository. For example, the Open Science Framework (https://osf.io) can serve this purpose.

Ideally, data are also deposited along with in relevant metadata, and in a repository that uses these metadata for indexing the data.

Finally, to make data *really* findable, these metadata should be semantic, but that is as yet out of reach for much of social sciences research (although, see this blog post for recent developments in this direction).

5 Open Materials

Publishing Open Materials is a bit different from publishing Open Data. On the one hand it's easier, because you will generally not risk publishing personal data as is the case for data. On the other hand it's harder, because although anonymized data generally exists in the public domain, materials you create are often creative works, and so by default, copyright applies. Finally, if you use closed infrastructure (for example, software that is not open source, like Qualtrics, NVivo, Atlas.ti, or SPSS), it is not always straightforward to make your materials open. You can publish your files, but the may require the prospective interested scientist to first pay for the same software, and that excludes quite a large proportion of scientists (i.e. all but the richest few).

Fortunately, all these topics are also relevant for other aspects of open science. You may have read about them already - and otherwise, see Chapter 8 and Chapter 2.

In any case you will need to choose a repository to deposit your materials. The Open Science Framework is generally a good choice. It can integrate with a number of third-party services to make the process as effortless as possible. You can create an account for free at https://osf.io.

5.1 Conferences

For conferences, organizers (or anybody, really) can request an OSF Meetings page. A number of examples are linked to from https://hpss.one. This allows people to upload slides and posters very easily: just by sending an e-mail.

A pro tip: e-mail your slides to the OSF meetings page a few days before the conference. They don't have to be done yet: you will still be able to change them later (the only thing you won't be able to change is the filename). OSF will create a new OSF repository for you with your slides.

Make sure that when you send the e-mail, you use one of the e-mail addresses associated to your OSF account. This way, you will be able to edit the repository that OSF creates later. The repository will still be linked to the OSF Meetings page. You can then replace the file with your slides or poster with updated versions (as long as the filename stays the same), and even add more files, or edit the repo's wiki to link to other interesting resources for your audience.

This new repo will be directly accessible from the OSF Meetings page of the conference, and like all OSF repositories, it will have its own short URL. You can then use this URL to create a QR code, and include the URL and/or QR code on your slides or poster. This

way, you can easily let people download your slides or poster (is using slides, make sure to include the URL and/or QR code on both the first and last slide, and explicitly tell people that it will be on the last slide - that way, people know that if they missed it, they can just wait until the end.

6 Preprints

A preprint is a version of a manuscript that is published before the so-called Version of Record (VoR) is published. The VoR is the version as it is produced by a publisher in a journal (generally with the publisher's layout etc).

6.1 Advantages of preprints

Preprints have a number of advantages. First, they can be published when (or before) you submit a manuscript, considerably speeding up science. Second, they are generally published Open Access, which makes your article open access even if it ends up being published behind a paywall (see Chapter 7). Third, they allow you to obtain peer reviews from more peers than the designated reviewers. Fourth, preprinted articles receive more citations than non-preprinted articles.

Because of this last reason, most editors (and so, journals) are in favour of preprinting. There are some exceptions though, and if you want to check, there is the excellent Sherpa Romeo website, which indexes all pubishers' and journals' policies regarding what you're allowed to do with which versions of the manuscript file: https://v2.sherpa.ac.uk/romeo/.

6.2 Using Open Preprint Servers

Also because of this last reason, many publishers now also created their own preprint servers, offering to preprint your article in the submission process. However, given that publisher's interests are not the same as those of science, it is better to not rely on their to provide this infrastructure. In addition, linking the preprint to the associated data, materials, and preregistration makes your products much easier to find by interested colleagues.

Therefore, using Open Science Framework preprint servers, such as PsyArXiv or one of the other preprint servers at https://osf.io/preprints is usually your best option. You can then immediately link the preprint to your OSF repository.

6.3 Linking the DOIs

Preprints receive DOIs so that they are instantly citeable. In addition, you can update the preprint with the DOI of the published version of the article. This allows bibliographic

databases and services such as OpenAlex to know that the two DOIs describe one academic work, which makes it possible to accumulate all citations .

6.4 Pro tips

If you post a preprint, you can add additional appendices at the end, and/or a frontpage where you link interested readers to additional resources. An example is available at https://doi.org/jnjp.

In addition, once your preprint's DOI is minted (i.e. its registration is processed by https://crossref.org, you can use CrossRef's ShortDOI service to create a ShortDOI. ShortDOIs always resolve to the corresponding DOI. ShortDOIs are convenient short unique identifiers for an article, great for e.g. slides. In addition, because the resulting URL is very short, the resulting QR code is simpler, allowing it to be scanned succesfully from a greater distance.

7 Open Access

Publishing Open Access means publishing something in a way that can be accessed for free by anybody. There are a number of ways to publish things Open Access.

7.1 Types of Open Access

First, if you publish preprints, you already publish Open Access: preprints are generally publicly available after all (see Chapter 6). You can update preprints, so that once your article is accepted, you can update it with the so-called "post-print version" of the manuscript: the version that was not typeset as it will be in the journal, but that does have all revisions implemented.

In addition, you can publish the Version of Record of an article Open Access. The Version of Record (or VoR) is the manuscript as it appears in the journal (including typesetting etc). There are also multiple ways to publish the VoR Open Access (or OA).

The way that is most consistent with Open Science principles is to publish in a Diamond Open Access journal (also known as a Platinum Open Access journal). Diamond OA journals do not charge so-called Article Processing Costs (APC). This is the price you pay to publish something OA in a Gold OA journal. At the bottom of this page, you find a list of Diamond Open Access journals in Psychology.

Gold OA journals also publish open access, but they charge authors to do so (the APC). This APC often includes substantial profit margins, and channeling public funds into corporate profit margins is ethically circumspect.

Before we go into ways to avoid this (apart from preprinting your articles and/or publishing in Diamond OA journals), there is one more type of OA journal: the hybrid journal. This is basically an old-fashioned closed access journal (also called a paywalled journal), where authors can opts to make an article OA on an article-by-article basis. Hybrid journals also charge APCs (in addition to the subscription costs they charge libraries or persons who want to read the paywalled content, essentially allowing them to double-dip and claim twice the public funds).

7.2 Gold for free

There are two ways to publish in Gold OA journals or Hybrid journals without paying the APC yourself. In the case of the first two ways, you don't even have to pay the APC at all.

The first way is to publish in a journal where your university has a publisher deal. For example, in the Netherlands, publishers negotiated publisher deals with a huge number of journals. The Wageningen Journal Browser allows you to find journals where authors from a given university can publish for free.

The second way works if you're in the Netherlands, and it leverages the Taverne Amendment. This allows you to publish open access if your work was funded by wholly or partly with Dutch public funds, even if you published it behind a paywall. For more information, see this site, and ask your librarians for local support.

7.3 Diamond Open Access journals in Psychology

- Biolinguistics: https://psychopen.eu/journals/bioling
- Clinical Psychology in Europe: https://psychopen.eu/journals/cpe
- Europe's Journal of Psychology: https://psychopen.eu/journals/ejop
- European Journal of Health Communication: https://ejhc.org
- Global Environmental Psychology: https://psychopen.eu/journals/gep
- International Journal of Open Educational Resources https://ijoer.scholasticahq.com/
- Interpersona: https://psychopen.eu/journals/ijpr
- Jahrbuch Musikpsychologie: https://psychopen.eu/journals/jbdgm
- Journal of Numerical Cognition: https://psychopen.eu/journals/jnc
- Journal of Open Source Education: https://jose.theoj.org/
- Journal of Open Source Software: https://joss.theoj.org/
- Journal of Social and Political Psychology: https://psychopen.eu/journals/jspp
- Measurement Instruments for the Social Sciences: https://psychopen.eu/journals/miss
- Methodology: https://meth.psychopen.eu
- Meta-Psychology: https://open.lnu.se/index.php/metapsychology
- Psychology of Human-Animal Intergroup Relations: https://psychopen.eu/journals/phair

- Quantitative & Computational Methods in Behavioral Sciences: https://psychopen.eu/journals/qcmb
- Research for People in Organizations: [https://psychopen.eu/journals/rpio(https://psychopen.eu/jour
- Sexual Offending: Theory, Research, and Prevention: https://psychopen.eu/journals/sotrap
- Social Psychological Bulletin: https://psychopen.eu/journals/spb
- The Quantitative Methods for Psychology: https://tqmp.org

Acknowledgement: Miriam Engels pointed out the 14 PsychOpen GOLD journals that were still missing, and Thomas Gültzow added the European Journal of Health Communication

8 Licenses and IP law

If you create something, by default it is copyrighted. That means that you own the intellectual property rights to it. This is true even if you created it in time you were employed with public money; and even if you endorse Open Science principles and so want to share the materials you create.

This copyright describes what others are allowed to do with your intellectual property unless or until you specify otherwise. What others are allowed to do pretty much boils down to: nothing. They are not allowed to copy your work, or sell it, or distribute it, or modify it.

Of course, this is the antithesis of Open Science principles, which mandate pretty much the opposite. Fortunately, being the intellectual owner of your works means you can override the default "rights" laid out in copyright.

You do this by attaching a license. Copyright itself can also be seen as a license, but one that disallows everything. To make it easy to specify what you want allow people to do with your work, the Creative Commons non-profit organization created a number of licenses ("CC licenses"). One of these is a different name for something that already existed: the public domain.

8.1 Public domain or CC0

The opposite of copyright is depositing a work in the public domain (also known as "CC0" or "Creative Commons Zero"). If you do this, you void all rights, allowing everybody to treat the work as if it were their own. This is the more liberal license: it prohibits nothing and allows everything.

8.2 The CC palette

In addition, you can choose a combination of rights from what you could call the CC palette. First, you can allow or disallow commercial use of your work. Second, you can allow or disallow modifications of your work. If you disallow modifications, others are only allows to copy or distribute your work unaltered. Third, if you allow modifications, you can implement a 'pay it forward' model where others who use your work must license the resulting work under the same license.

After applying a CC license, you're *still* allowed to given individuals or organizations extra rights. For example, if you apply a CC-NC license ("Non Commercial"), you can still choose to give an individual or organization rights to sell your work.

To look at all Creative Commons licenses, see https://creativecommons.org.

8.3 Data licenses

For data, a specific set of licenses has been created, that are similarly easy to understand and apply as the Creative Commons licenses. These were created by Open Data Commons (see the pattern here?), and you can view them at https://opendatacommons.org.

8.4 Software licenses

For software, licenses are considerably more complicated. There's an overview of what you are and aren't allows to do with the most popular ones at https://choosealicense.com/licenses, and there is a more extensive overview at https://opensource.org/licenses.

One license that's a good default if you don't want to delve into this is the Community Research and Academic Programming License (the CRAPL), developed by Matt Might. Its premise is simple:

> If you're an academic, the CRAPL is an excuse to convert ugly software into good karma.
>
> Matt Might

You can learn more about the CRAPL at https://matt.might.net/articles/crapl/.

9 Open Educational Resources

Open Educational Resources, or OERs as they are commonly known, are resources for educating yourself or others that are available under permissing licenses, such as CC-BY (see Chapter 8).

Producing OERs, therefore, is as simple as applying such a permissive license. Many OERs already exist, so you may also want to re-use an existing resources instead. Similarly, if you want to produce an OER, you may want to consider how to make sure people can find it. One way of doing so is exploring whether you may want to contribute chapters to existing open access textbooks, or submit your contribution to a diamond open access journal that accepts didactic materials such as tutorials.

To get you started, and if you happen to be a psychologist (such as the authors of this guide), this is a list of open access resources:

- The Noba Project collects chapters that you can then combine into a textbook: https://nobaproject.com
- The Open Textbook Library lists open access psychology textbooks: https://open.umn.edu/opentextbooks/subjects/psychology
- The second edition of the Psychology textbook at OpenSpax: https://openstax.org/details/books/psychology-2e
- Open Education Resources Commons has many OERs: https://oercommons.org/
- For more advanced topics, there the Diamond Open Access journal The Quantitative Methods for Psychology has useful articles: https://tqmp.org

10 Workshops and trainings

10.1 ARPH 2024, Maastricht (~90 minutes)

The slides for the presentation part are here.

10.2 Psychology faculty of the Open University of the Netherlands, 2024, Utrecht (~60 minutes)

The slides for the presentation part are here.

References

UNESCO. 2021. “UNESCO Recommendation on Open Science. Programme and Meeting Document SC-PCB-SPP/2021/OS/UROS.” UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000379949.