# Open Science Essentials

**Ein Überblick zu offener Forschung mit ausgewählten Hands-on Vertiefungen**

Jürgen Schneider ju.schneider@dipf.de

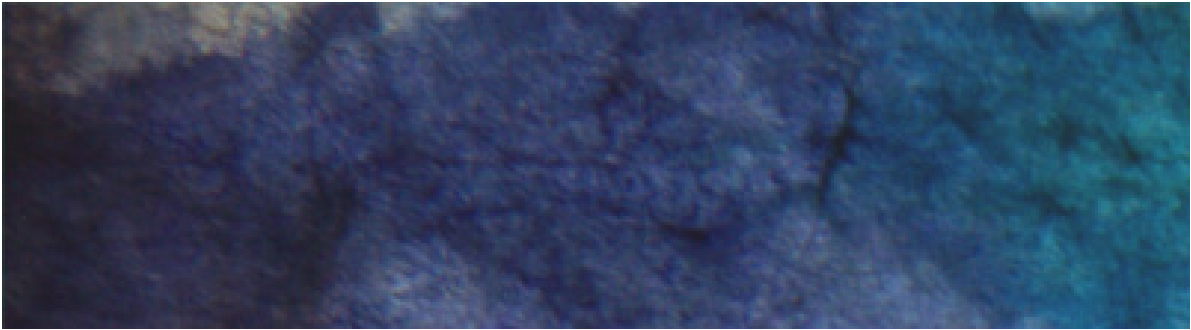17 January 2024

# Table of contents

# Overview

This is the overview to the workshop. Here you'll find the...



1. Introductory slides
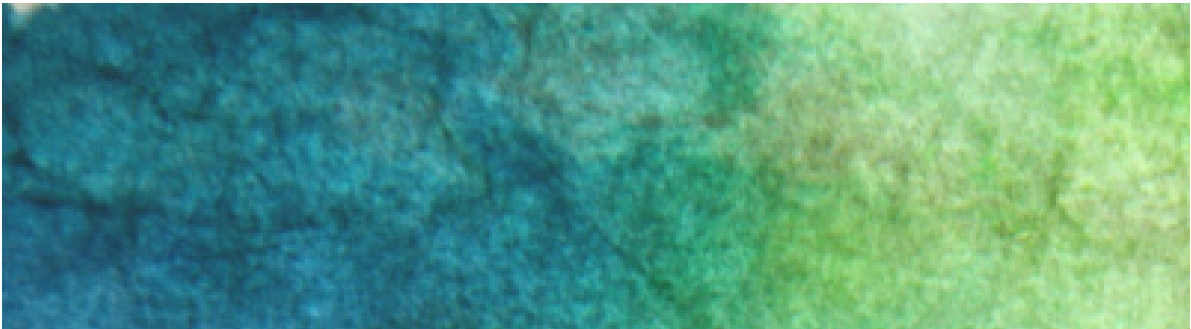


2. Tasks onopen access
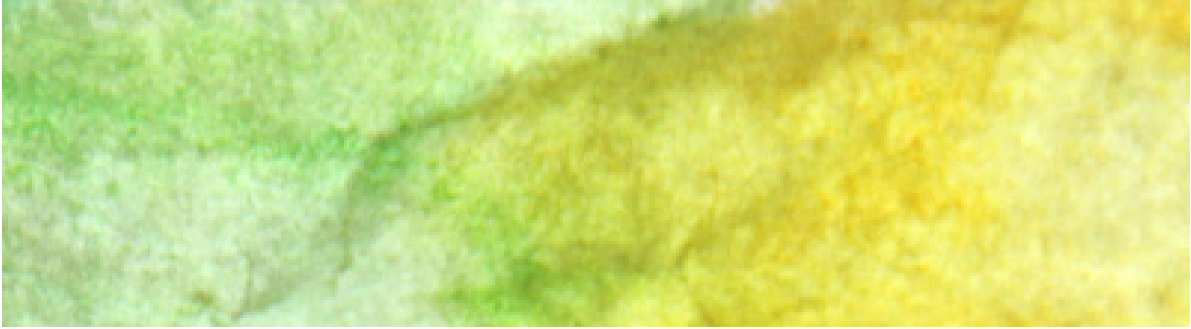
3. Tasks onopen data



4. Tasks on transparent inquiry



5. Closingslides

**Rough timeline**

- Introduction (50min)
  - Break (15min)
- Hands-on (50min)

- – Break (15min)
- Hands-on discussion (20min)
- Reflection (30min)

*Github Repo of this workshop, see [github.com/j-5chneider/OS-PHKA](github.com/j-5chneider/OS-PHKA)*

# Part I

# Open Access

In this chapter you will learn about

1. What is open access to scholarly publications?
2. What options do you have?
3. What are easy steps to get started?

# 1 What it is

Open access refers to **free access** to digital scientific publications via the internet that is not hindered by payment barriers or registration hurdles.
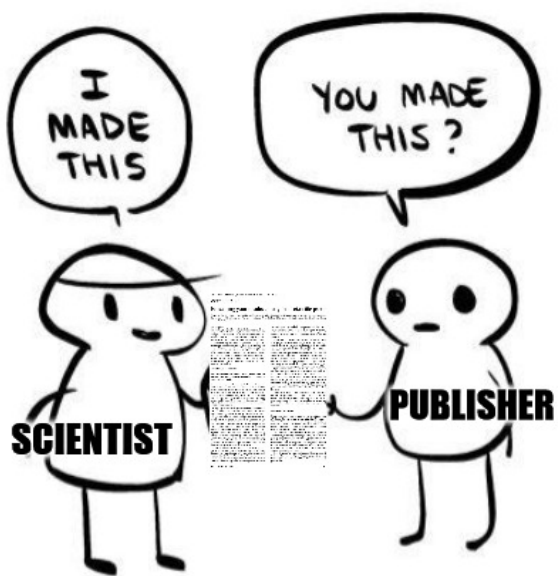
The rights of use for open access publications are regulated transparently via so-called free licenses (e.g. Creative Commons licenses).

The authors' copyrights, such as the protection of intellectual property or the right to quote, are also protected in open access publications in accordance with the Copyright Act.

In some (mostly closed-access) journals, the copyright is transferred to the journals and **the authors retain very few rights**. For example, if they want to use a graphic from the paper in a presentation, they would (theoretically) first have to request the right of use from the journal.

# 2 Options

(Severin et al., 2018)



## Diamond (= Platinum)

This model involves journals that are entirely free to readers and authors. The costs are usually covered by institutions, societies, or other funding sources.

*Example journal:* [Meta-Psychology](Meta-Psychology)



## Gold

In this model, articles are freely accessible to the public immediately upon publication.

The costs, so called **APC**s (article processing charges, see below) associated with publishing are typically covered by the author, their institution, or funding agencies.

APCs have to be paid only after the manuscript is accepted. Any other practice, like paying to initiate the review process is predatory publishing (see below).

*Example journal: AERA Open (moderate APC)*
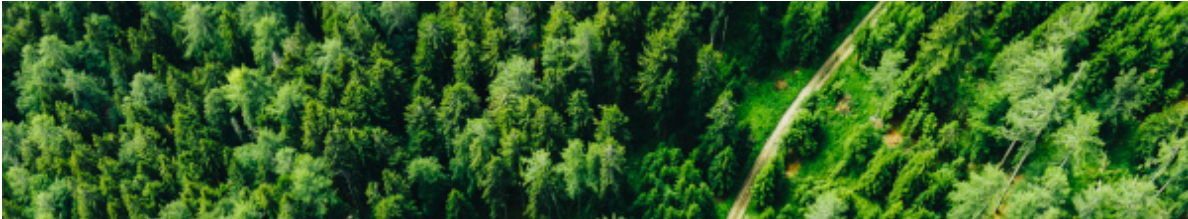
## Green

Also known as "self-archiving" of **pre-prints**. This approach allows authors to deposit a version of their article in an open **repository**. The archiving has to be done by yourself.

Some journals impose an **embargo period** on articles published with them ( *boo!* ). The pre-print cannot be shared before this period has expired.

*Trusted repositories: zenodo, SocArXiv*
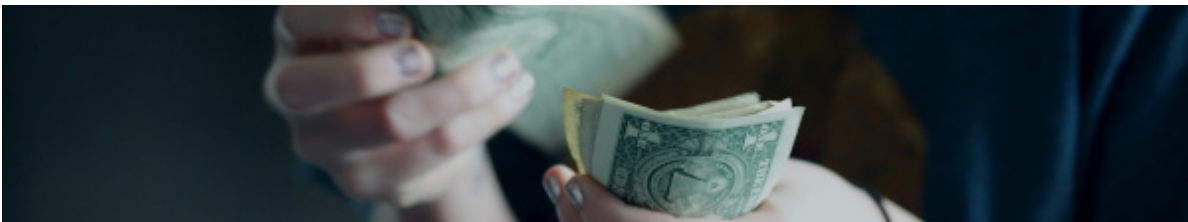
## Hybrid

Journals based on this model are closed, but offer the possibility to **"unlock" your individual article**, which then becomes open access. APCs for hybrid journals are usually outrageously expensive ( *boo!* ).

Since all the other articles in the journal may still be behind a paywall, libraries may still have to pay a lot of money, as they usually buy entire issues with several articles. In this way,

taxpayers pay twice for the process of opening the access to the paper.

*Example journal:* *Computers and Education (not so moderate APC)*

---

**🔍 Task**

- Google search for a journal article you recently read and find it on the publisher's webpage
- Which model of open access does this journal use (gold, diamond, hybrid, closed)?
- Imagine you would submit a manuscript to this journal: How much would it cost you to make your article open access?

---

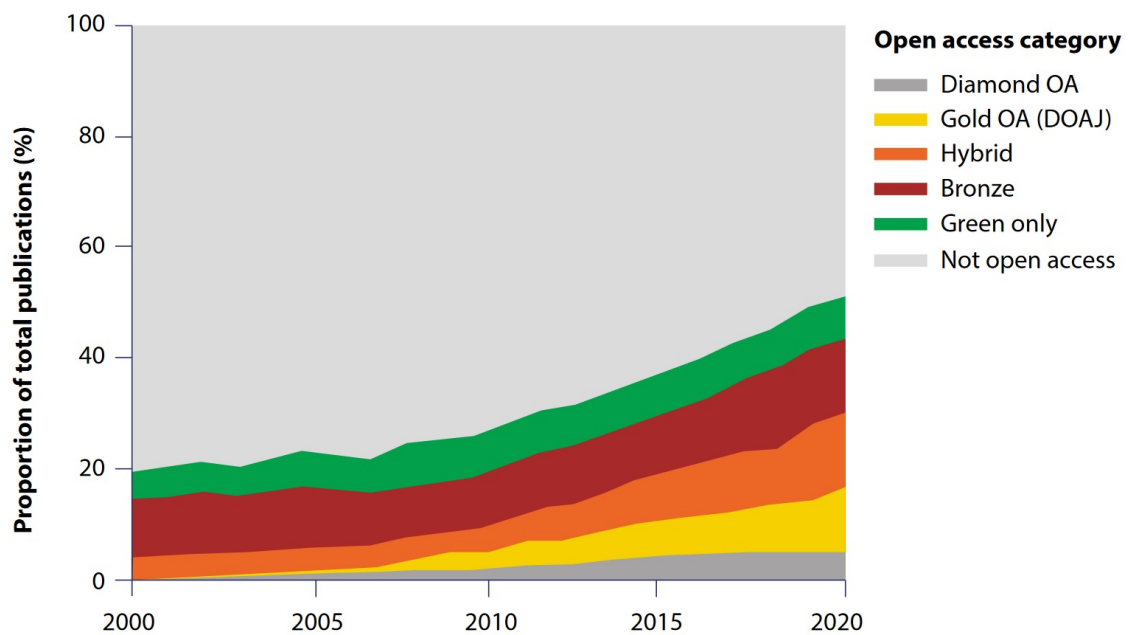**more info on...**

How is the distribution?

*Fig 1. Share of scientific publications by access type in their year of publication (bronze open access = without license)* (UNESCO, 2023)

## APC

**Where do I get the money for the APC?**

- If the project is supported by a research funder, open access publication costs are usually included as part of the grant (or can be applied for).
- Research institutes typically have an open access publication fund. Check at the library of your institute. However, the eligibility of the journal for the fund is subject to certain conditions, which are defined in the terms of use of the open access fund or the open access policy (if available). For example, publications in hybrid journals are often not getting funded.

## Predatory Publishing

**Some publishers**
engage in fraudulent business practices.

A typical way is that such journals promise fast review procedures and publication times. These are indeed fast, but because they lack proper editorial standards and peer review processes. There is no real critical appraisal of the manuscript. Therefore acceptance rates are very high and publication can be very fast. Also, authors often have to pay an APC already *before* the initiation of the review process.

The website thinkchecksubmit.org helps you to sort out the good from the bad.

**Some researchers and librarians**
are also critical of the so-called "mega-journals". (Oviedo-García, 2021)

These include for example PLOS ONE and the Frontiers series. The business model of these journals is based on the fact that they publish a *very* large number of articles and collect APCs for them. The criticism is that the editors place their priority on a high publication output, with the quality of the review process suffering as a result.

The rapidly increasing number of special issues in regular journals is also viewed critically for the same reasons.

## Preprint

A preprint is a **version** of a scientific manuscript **before it is submitted to peer review** and publication in a scientific journal. At this stage, publishers or peer reviewers have not yet invested any resources in the manuscript, which is why the sole intellectual property rights lie with the authors.

This means: Please *save different versions* of your manuscript! Either by yourself of using version control software (like git). Otherwise, it will be difficult to restore the version as it was before you submitted it.

Preprints can be uploaded to so called "preprint servers" and repositories free of charge. Typical repositories in education: zenodo, SocArXiv*

**When to upload**

- Pre-submission (actual pre-print): Some researchers choose to upload a preprint before formally submitting their manuscript to a journal. This allows for early dissemination of findings and the opportunity to receive feedback from the community before the formal peer review process begins.
- Post-publication: Even after a manuscript is formally published, researchers may choose to upload a preprint to maintain an openly accessible version of their work, ensuring broader dissemination.

*In both cases please check with the journal you are about to submit to, if this is in line with their publishing guidelines!*

# 3 Easy first steps

**🔍 Finding open access journals**

A good first glimpse at what are good options for open access journals is [doaj.org](doaj.org). The Directory of Open Access Journals (DOAJ) systematizes information on journals that are open access.

- Go to [doaj.org](doaj.org)
- Type in a keyword that describes your research (e.g. "digital", "school", …)
- Refine your results by applying the filters on the left (e.g. subjects, languages, …)
- Can you find interesting journals? Explore the results and their

  - APC
  - publisher
  - expected time from submission to publication
  - …

- Note interesting journals to share with your colleagues

**🔍 My first preprint**

Sharing preprints is incredibly easy. Let's try it!
If you don't have a manuscript to share, just use a blank PDF page (e.g. by saving one from MS Word).

- Take your PDF and head to [SocArXiv](SocArXiv) or [Zenodo](Zenodo)
- Upload your PDF
- Explore and compare the features between SocArXiv and Zenodo:

  - How can you link the preprint to your article in the journal (if it is already published)?
  - How can you upload a new version of the preprint?
  - How can you restrict access to the preprint or apply an embaro period?

- What information (= metadata) can you provide about your preprint?

- If you would like to share a preprint: Which platform would you choose?

# Part II

# Open Data

In this chapter you will learn about

1. What is open data?
2. What options do you have?
3. What are easy steps to get started?

PUBLICATIONS AND DATA

# 4 What it is

Research data is open, if

- **anyone**
- can **readily access** the data
- at no more than a **reasonable reproduction cost** (i.e., internet connection)

(Open Knowledge Foundation, 2023)

> **i** Openness is not a dogma and not a dichotomy
>
> There may be perfectly good reasons not to share data or to share them only with a very specific group. Such as the protection of vulnerable or marginalized individuals or due to ministry regulations.
>
> This is why the general notion when sharing data is:
> *"As open as possible as closed as necessary"*
> (European Commission, 2023, p. 36)

# 5 Options

Depending on whether you are sharing data for reasons of **transparency** or to enable **reuse**, a different level of preparation is required and a different archiving facility might be considered.

> 💡 Is my data fit for reuse?
>
> Here are a couple of guidelines to decide, if your data is fit for reuse
>
> - Costs:
>
>   - How much **effort** is required for well-documented data sharing (e.g., does a codebook exist? What steps are necessary for data cleaning/editing?)
>   - Is there **consent for reuse** available or would it have to be obtained retrospectively?
>
> - Benefits:
>
>   - How large is the remaining **analysis potential**? (e.g., not fully analyzed, type of data, connected with other data sources)
>   - How high do you rate the **quality** of the data? (e.g., representative, size, special features of sample)

### Share for transparency

Put you data into a **repository**!
Recommended options are:

- Zenodo
- psycharchives.org
- osf.io

### Share for reuse

Put you data into a **research data center**!
Recommended options are:

- Verbund FDB (Education)
- RDC at ZPID (Psychology)
- re3data.org (Database to search for databases)

### I can't share

**Archive** your data, so you'll be able to access it **long-term**!
Recommended options are:

- Repository of your institution (if available)
- Using a public repository (e.g., Zenodo, osf.io) and switch status to "private"

## 5.1 FAIR data

Researchers and infrastructure specialists realized: Simply putting a lot of data out there is **not effective**. Making data FAIR (findable, accessible, interoperable, reusable) aims to mitigate this problem (Wilkinson et al., 2016).

go-fair.org

### Findable

**The problem:**
Just because we provide data online, doesn't mean that others will find it.

We could have the greatest data set to answer further research questions - if our colleagues don't know it exists or can't locate the data, openness will be of little value.

**The solutions:**

- Get a persistent identifier (e.g., DOI), where you provided your data

  - search for a research data center that fits your needs: re3data.org
  - recommended research data centers: Verbund FDB (education, Germany), RDC at ZPID (psychology, Germany), …
  - recommended repositories: Zenodo, psycharchives.org, osf.io, …

- Mention DOI in publication that builds on this data (e.g., in the "data accessibility statement")
- Describe your data as richly as possible (metadata). *Research data centers* offer form fields tailored to the discipline or data type. With *repositories* use alternative possibilities, such as keyword fields.

  - e.g., which variables does the quantitative data set contain?
  - e.g., which topics does your data cover?
  - e.g., which population did you draw your sample from?

### Accessible

**The problem:**
Just because others find our data doesn't mean the *access barriers* are as low as possible and doesn't mean they know *in which way* they are allowed to access it. Examples:

- Providing a link to the data in the text of a paywalled journal article

- Unclear licensing / use conditions when providing data (e.g., are non-researchers allowed to access the data or is it only open for qualified researchers?)

**The solutions:**

- Make sure access is free of charge (or as cheap as possible)

  - e.g., by providing link to data in publicly accessible sections of journal articles that are not open access
  - e.g., by using repositories or research data centers that allow access free of charge

- Make sure users know if they can access and under which conditions

  - e.g., *research data centers* ensure that terms of use are clear (who may access under what conditions) and offer different levels of access restriction
  - e.g., on *repositories* provide a readme-file and an open license (e.g., CC0, CC-BY, CC-BY-SA) with data sets for access cases

## Interoperable

**The problem:**
Just because others downloaded our data doesn't mean they can open and manipulate it.

**The solutions:**

- Use file formats with open licenses

  - e.g., tabular data: CSV (with additional labelling script), RData
  - e.g., text data: PDF, HTML, ODT, RTF

- Make sure users know how different files are related to one another

  - e.g., define which file contains student data and which teacher data
  - e.g., define which file contains data from cohort 1 and which cohort 2, …

## Reusable

**The problem:**
Just because others opened our data doesn't mean they understand the data and its use-conditions. Examples:

- Others can't understand what the column names of the tabular data set mean: Which columns in the data set relate to which variables in the journal article?
- Can someone from sociology use the data set from psychology they found on osf.io?
- Does someone reusing a data set have to cite the authors?

**The solutions:**

- Adhere to standards in folder organization

    – e.g., PSYCH-DS (see technical specification draft)

- Rich description/explanation of what user will find *in* the data set ( meta descriptions about the data set *as a whole*, as for accessibility)

    – e.g., provide a codebook. How to semi-automatically create a codebook, see the R package `codebook`

- Provide a license for the use-cases

    – again, *research data centers* ensure that terms of use are clear (who may use under what conditions)
    – again, on *repositories* provide a readme-file and an open license (e.g., CC0, CC-BY, CC-BY-SA) with data sets for the use-cases

---

**ℹ FAIRness vs. openness**

"does not necessarily mean that data has to be "open" [...] even highly protected data can be FAIR data"
(Kraft, 2023)

---

# 6 Easy first steps

**Choose one of these topics:**

(And if you are done maybe do the other. Or do as you like - I don't make the rules.)

> **Q** Check the informed consent
>
> If your data includes personal information, the European GDPR **requires a declaration of consent** from the respondent before it is shared.
>
> - Check the consent form of your study. Does it cover data sharing (in the *purpose of the study* section)?
> - A typical option is to obtain *"broad consent"*. This includes not only the processing of data for the purposes of your study, but also for other studies.
>   - Check out the template for informed consent (VerbundFDB, 2018), particularly under "weitere Nutzung" on page 4 and compare it to your consent form.
>   - In which case do you *not* need the participants' consent to share the data?
>   - (Additional: Does your consent form check all the boxes of the checklist of legally compliant consent forms? (VerbundFDB, 2019)

> **Q** Let's explore repositories for data sharing!
>
> **How suited** are these repositories for sharing data **publicly** (for transparency) or for archiving data **privately** (because you can't share)?
>
> - Each of you go to one repository: Zenodo, psycharchives.org, osf.io
>   - Upload a random CSV file.

- – Is there an option to make your repository private?
- – Is there a possibility to set an embargo period, in case you want to wait until your paper got published?
- – How can you link your publication to the data and your data to the publication?
- – Can you associate a license with your files, so users know under which conditions they can use it?
- – Let's assume you have a codebook for your data set. How can you associate it with the data?

- Report back to each other what you found.

# Part III

# Transparent inquiry

In this chapter you will learn about

1. What is transparent inquiry?
2. What options do you have?
3. What are easy steps to get started?

# 7 What it is

In qualitative research, one of the key challenges is to make the epistemological process, i.e. the interpretations and conclusions drawn from the material, **intersubjectively comprehensible**. In conventional publication formats, the sources or data cited and the **analysis process on which the authors base their conclusions** are usually only partially transparent. (Kapiszewski & Karcher, 2021)
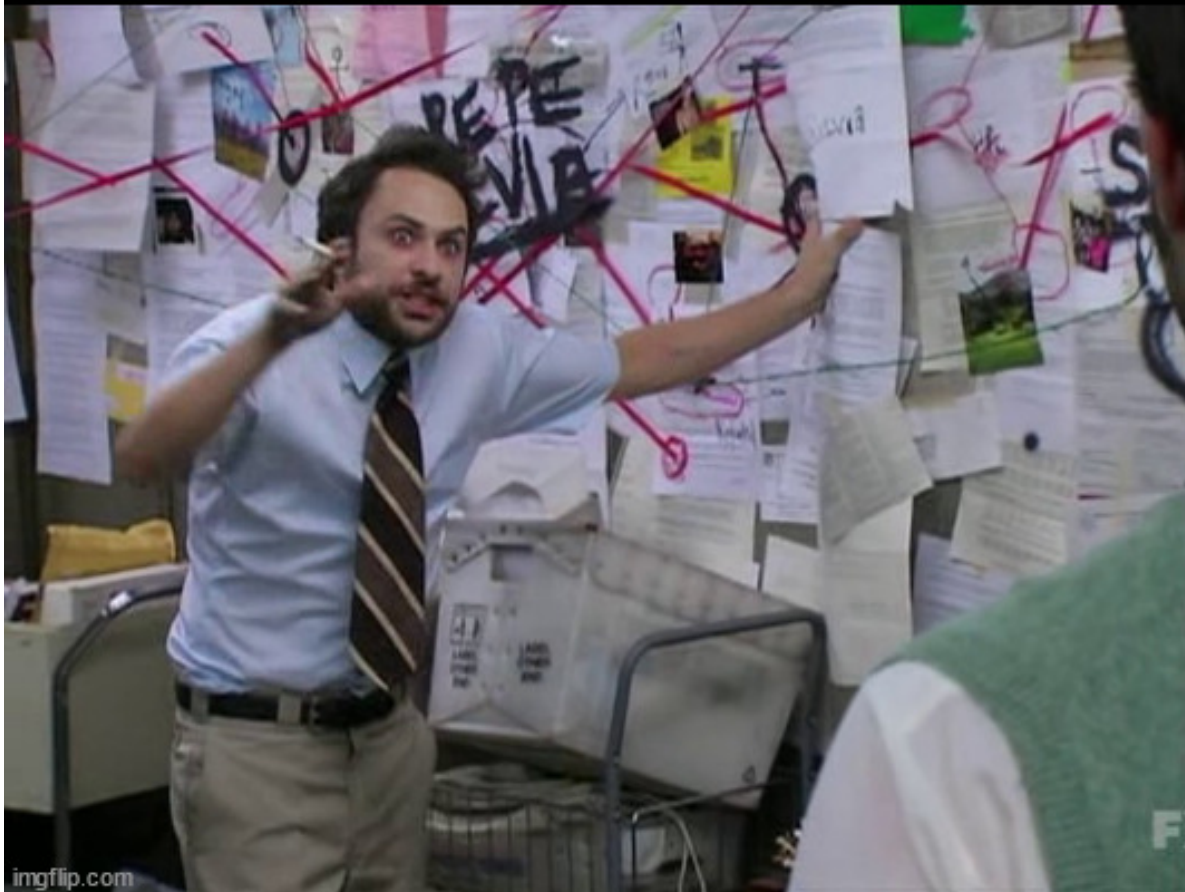
Transparent inquiry is the attempt to highlight parts of the text in the manuscript and link these to sections of the data, descriptions of the approach used in the analysis, interpretations and conclusions.

The **basis of the argumentation**, i.e. the way in which the authors' conclusions lead to the scientific claims in the manuscript, is thus directly linked to the scientific claims themselves, making the reasoning tangible to the reader.

# 8 Options

There are currently **few established standards** and workflows for this challenge that are capable of achieving consensus in a broader community. (Humphreys et al., 2021)

## Annotation for Transparent Inquiry (ATI)

... is a tool that has already been piloted and explored, for which initial experience reports and publications using the tool are available.

[qdr.syr.edu/ati](qdr.syr.edu/ati)

ATI is an open **web annotation tool** that allows annotations to be generated and shared over the web (Elman & Kapiszewski, 2018). Similar to the commenting function in word processing programs, authors can mark parts of the text in the manuscript and link them to excerpts from the data, descriptions of the analysis procedure, interpretations and conclusions.
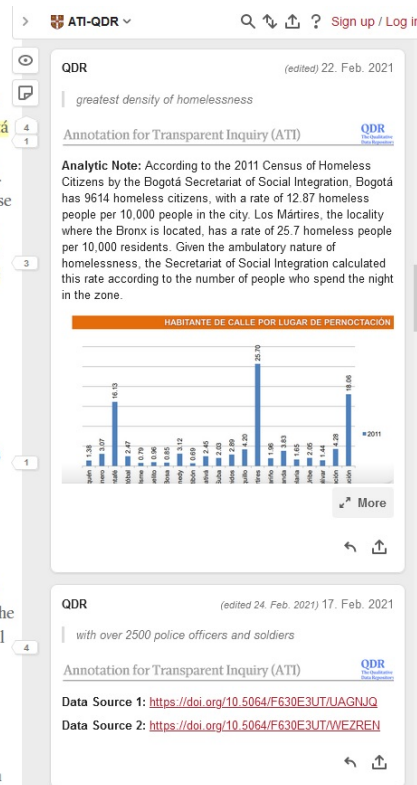
ⓘ Example of a manuscript annotated with ATI

Figure: Example of a manuscript annotated with ATI (Mayka, 2021)

The tool uses the open source annotation software Hypothes.is (www.hypothes.is) and therefore works online web text as well as PDF documents (such as on repositories, e.g., SocArXiv). Consequently, **every** digitally published **article** and every excerpt from it **can be annotated**.

While the annotations are created by the authors, they are hosted and curated by the Qualitative Data Repository (www.qdr.org) of Syracuse University. This makes the annotations machine- and human-readable in the **long term**.

# 9 Easy first steps

🔍 Explore the use of ATI in articles

There are several **examples** that use ATI listed on its website: qdr.syr.edu/ati/ati-models

- Select one or two of the articles from the table and open the annotated version of the article (link in column "ATI Link").
- What is your first impression on the intersubjective comprehensibility? Does ATI improve your understanding of the author's reasoning?
- How did different authors use ATI in their articles?

🔍 Let's try out hypothes.is!

For exploration purposes, we'll try to **annotate an article** of someone else.

- Go to (web.hypothes.is/start) and follow the steps there:
  - Make an account
  - Install the Chrome-extension or save the Firefox/Safari bookmarklet
- Now let's annotate an article:
  - Head to this paper doi.org/10.3389/feduc.2023.1166018
  - Activate hypothes.is (will show on the very right of your screen)
  - Explore how you can annotate (text, pictures, links, …)

# References

Elman, C., & Kapiszewski, D. (2018). The Qualitative Data Repository's Annotation for Transparent Inquiry (ATI) Initiative. *PS: Political Science & Politics*, *51*(01), 3–6. https://doi.org/10.1017/S1049096517001755

European Commission. (2023). *Horizon Europe (HORIZON). HE Programme Guide. Version 4.0.* Publications Office.

Humphreys, L., Lewis, N. A., Sender, K., & Won, A. S. (2021). Integrating Qualitative Methods and Open Science: Five Principles for More Trustworthy Research*. *Journal of Communication*, jqab026. https://doi.org/10.1093/joc/jqab026

Kapiszewski, D., & Karcher, S. (2021). Transparency in Practice in Qualitative Research. *PS: Political Science & Politics*, *54*(2), 285–291. https://doi.org/10.1017/S1049096520000955

Kraft, A. (2023). *The FAIR Data Principles.* https://doi.org/10.23668/PSYCHARCHIVES.13577

Open Knowledge Foundation. (2023). What is Open Data? In *Open Data Handbook.* https://opendatahandbook.org/guide/en/what-is-open-data/.

Oviedo-García, M. Á. (2021). Journal citation reports and the definition of a predatory journal: The case of the Multidisciplinary Digital Publishing Institute (MDPI). *Research Evaluation*, *30*(3), 405–419a. https://doi.org/10.1093/reseval/rvab020

Severin, A., Egger, M., Eve, M. P., & Hürlimann, D. (2018). Discipline-specific open access publishing practices and barriers to change: An evidence-based review. *F1000Research*, *7*, 1925. https://doi.org/10.12688/f1000research.17328.1

UNESCO. (2023). *Open science outlook 1: Status and trends around the world.* UNESCO. https://doi.org/10.54677/GIIC6829

Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., … Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, *3*(1), 160018. https://doi.org/10.1038/sdata.2016.18

# Credit

Overview: Alexander Grey on Unsplash

Diamond: Edgar Soto on Unsplash

Gold: Jingming Pan on Unsplash

Green: Miha Rekar on Unsplash

Hybrid: Alexander Grey on Unsplash

Publications and data: CC-BY aukeherrema.nl

Transparency: Max Harlynking on Unsplash

Reuse: Thomas Lefebvre on Unsplash

Can't Share: British Library on Unsplash

Icons: CC BY 4.0 Fontwaesome