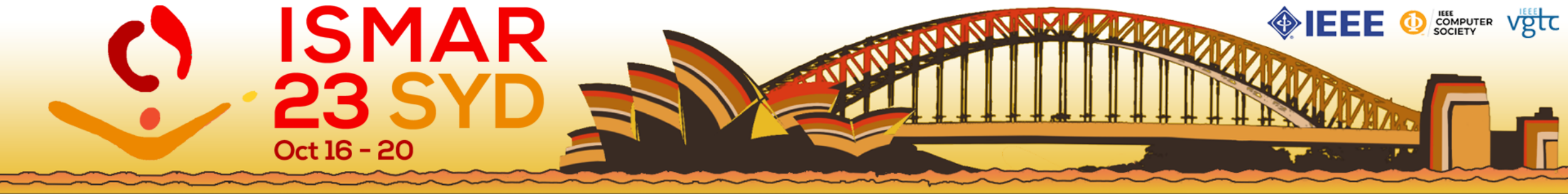
Replicability and data/algorithm storage and availability

3rd Workshop on Replication in Extended Reality (WoRXR)

Florian Weidner

Lancaster University



Outline for the next 15 minutes

- **Why** publish data and algorithms?
- **What** could (should) we store?
- **Where** to store?
  - Zenodo
  - OSF
- How do we **document** it?
- Summary

Why publish data?

Reproducibility and replicability. For example:

- Prototype shared: test exactly the same.
- Data shared: calculate exactly the same
- Everything: do exactly the same.

But also:

- Comparison with new findings easier (statistics)
- Making improvements is easier (iterations)

What could (should?) we store?

- *Simulation data*
- *Data from surveys, questionnaires and interviews*
- *Audio-visual data*
- *Software, code, scripts*
- *Measurement data (raw + processed)*
- *Analysis data*
- *...etc.*

Where to store?

- „Popular“: osf.io, zenodo.org, IEEEDataPort
- Some institutions have platforms
- Criteria:
  - Offers DOI
  - promises long-term storage
  - Stable funding
  - (Free)

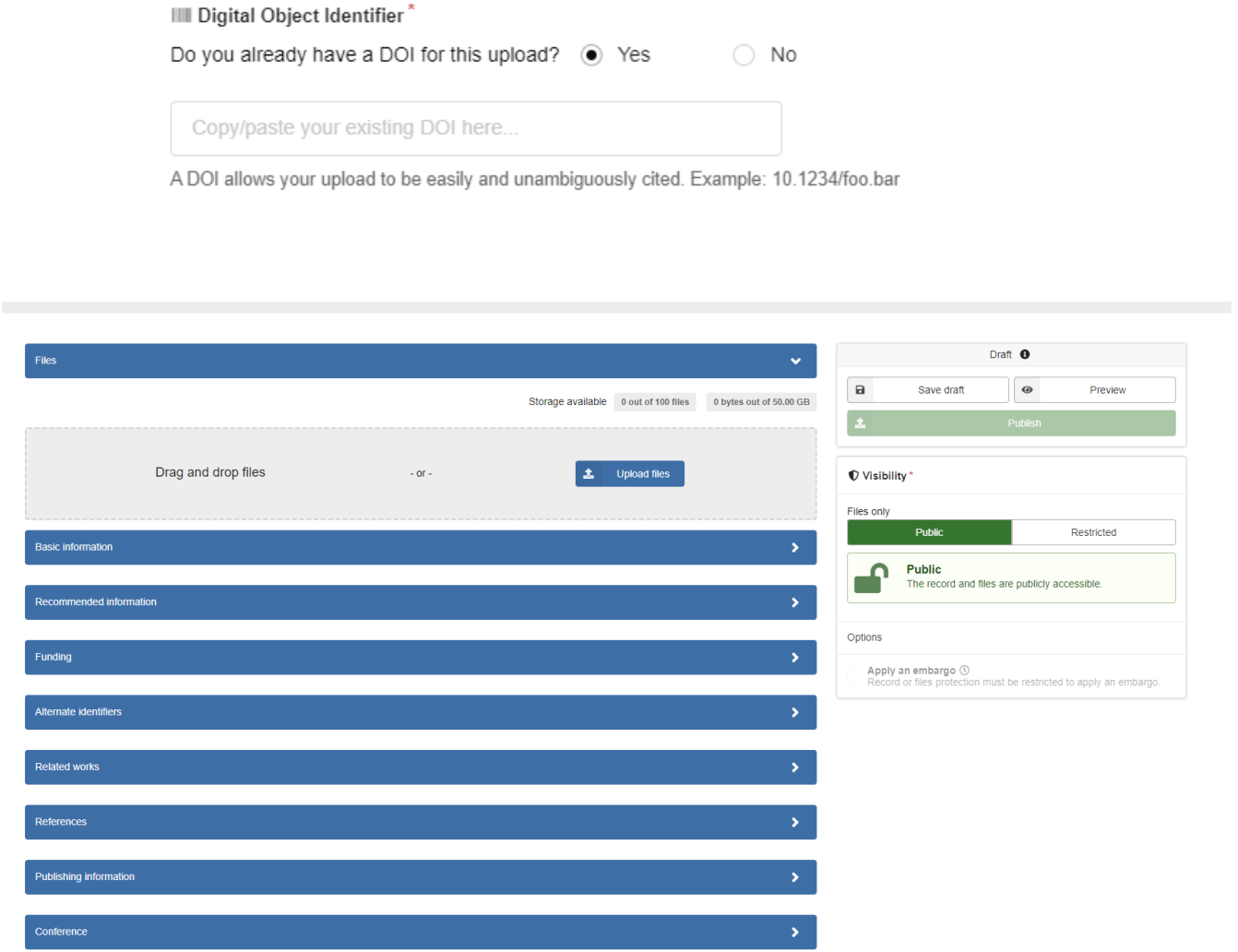
Zenodo

All about storage.

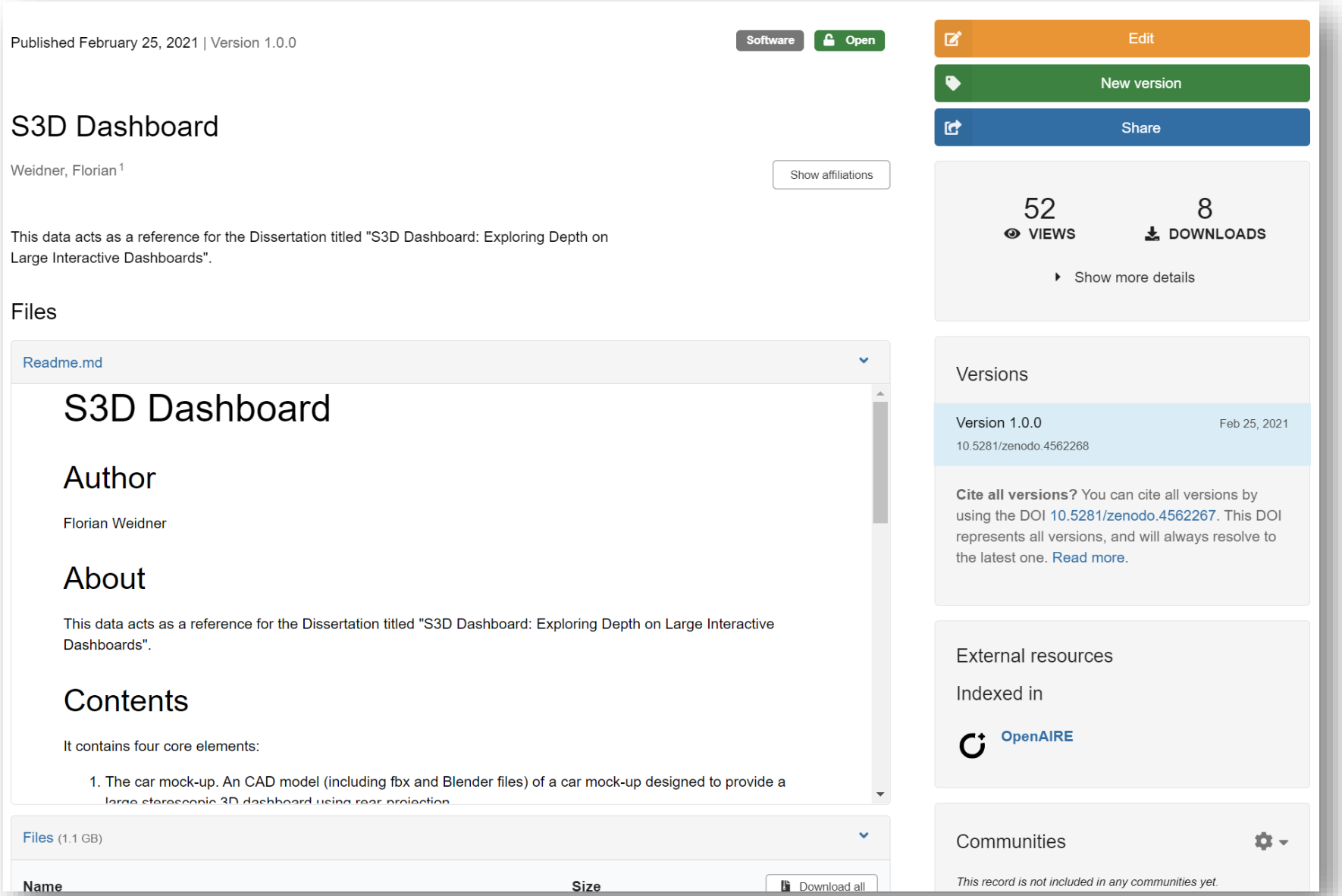
Zenodo UI

- Signup / login
- Create new upload
- Get DOI
- Fill in metadata
- Upload files
- Review
- Publish

➔ straightfoward



Example



Published February 25, 2021 | Version 1.0.0

Software Open

S3D Dashboard

Weidner, Florian [1]

Show affiliations

This data acts as a reference for the Dissertation titled "S3D Dashboard: Exploring Depth on Large Interactive Dashboards".

Files

Readme.md

S3D Dashboard

Author

Florian Weidner

About

This data acts as a reference for the Dissertation titled "S3D Dashboard: Exploring Depth on Large Interactive Dashboards".

Contents

It contains four core elements:

1. The car mock-up. An CAD model (including fbx and Blender files) of a car mock-up designed to provide a large stereoscopic 3D dashboard using rear projection

Files (1.1 GB)

Name Size Download all

Edit

New version

Share

52 VIEWS

8 DOWNLOADS

Show more details

Versions

Version 1.0.0 Feb 25, 2021

10.5281/zenodo.4562268

Cite all versions? You can cite all versions by using the DOI 10.5281/zenodo.4562267. This DOI represents all versions, and will always resolve to the latest one. Read more.

External resources

Indexed in

OpenAIRE

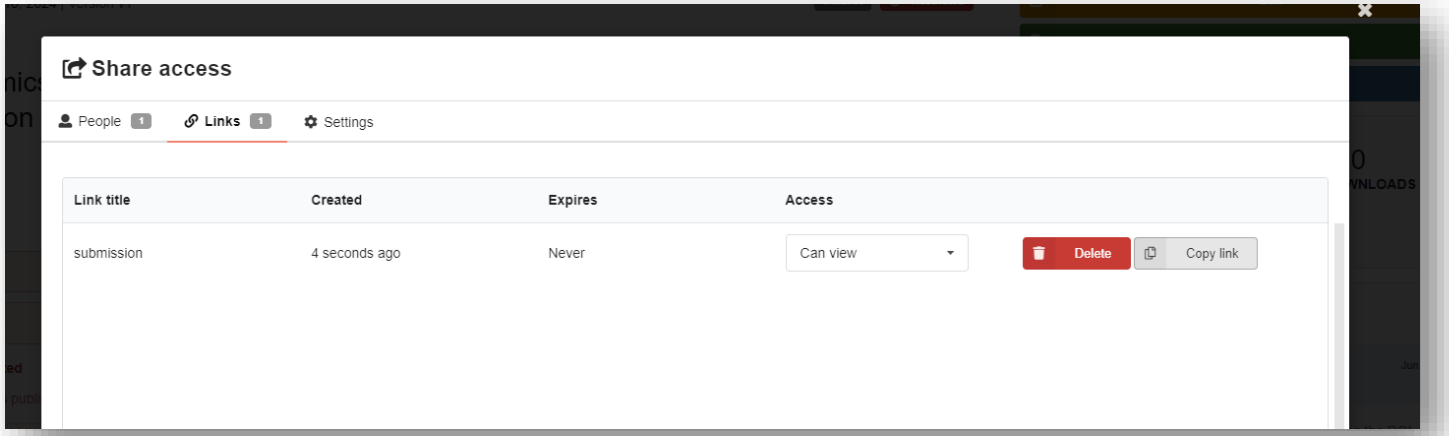
Communities

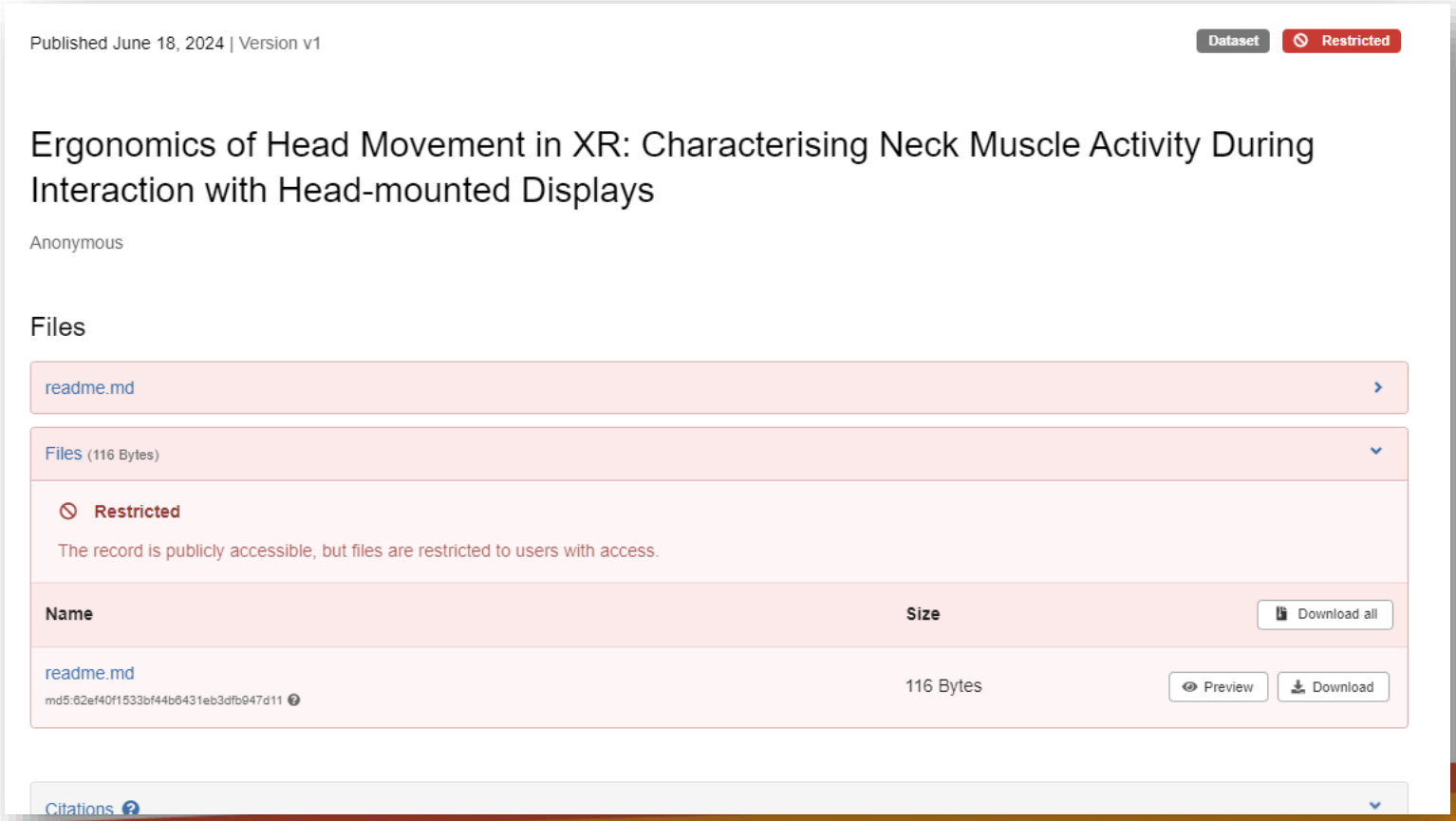
This record is not included in any communities yet.

Private?

1. Create anon repository
2. Publish restricted
3. Create link
4. Share link

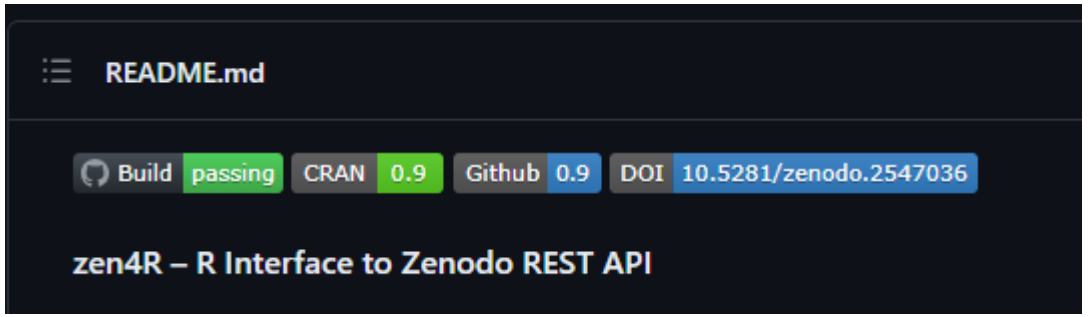
(change authors later)





Comments on Zenodo

- Free
- Funded by CERN (long-term perspective)
- No true anonymous links;collaborators can be modified
- Versioning possible
- 50GB
- Links to github release
  - auto-update on new release
  - Nice github Zenodo badge



https://github.com/eblondel/zen4R

OSF (Open Science Framework)

It’s all about (pre-)registration, project, and resources: lifecycle

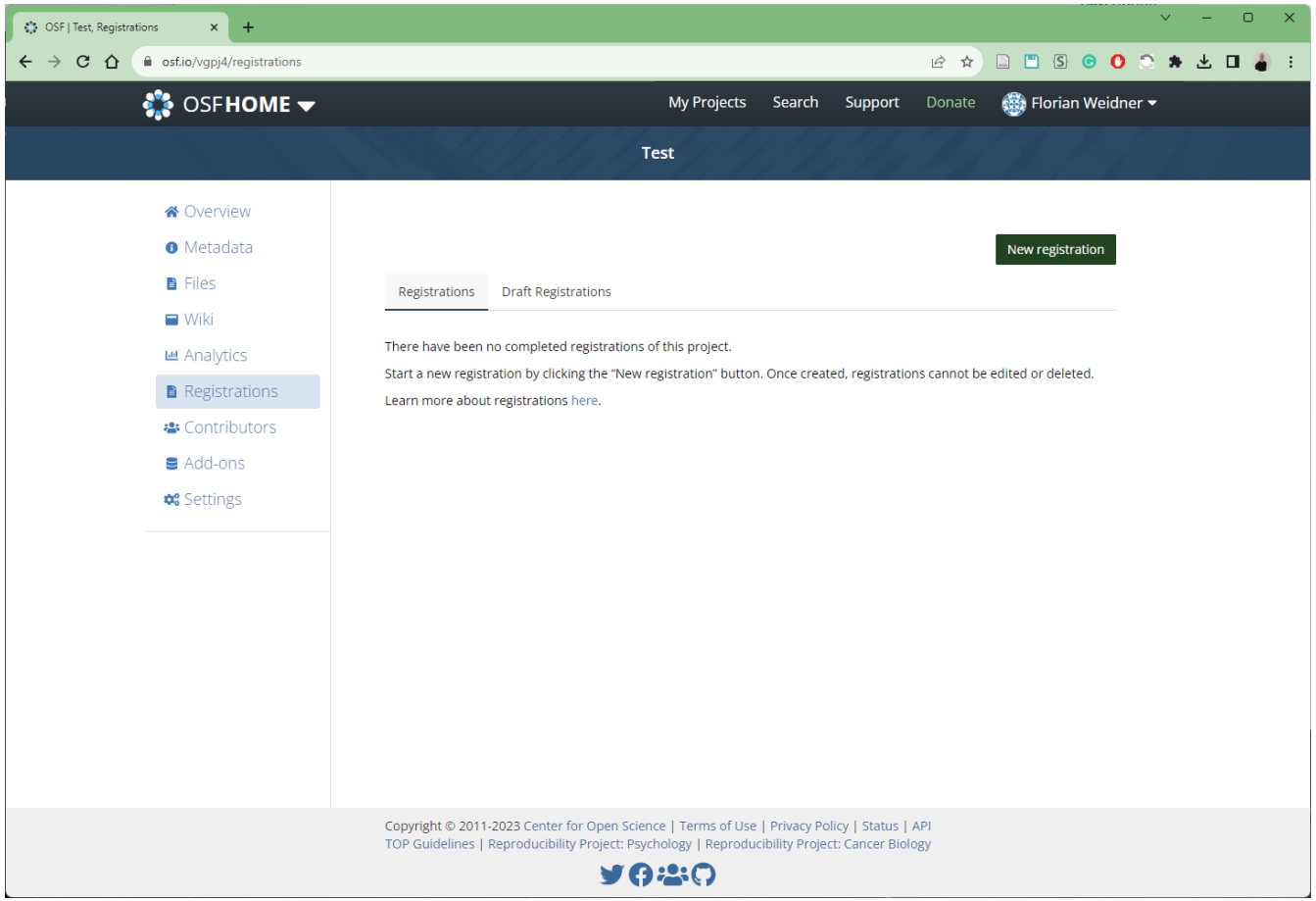
TIL: “OSF is maintained and developed by the Center For Open Science (COS)”

Registration + Project (+ Preprint)

- Registration is the start
- Registration creates project
- Project is for data and collaboration
  - Is flexible
  - Own storage or external
  - History

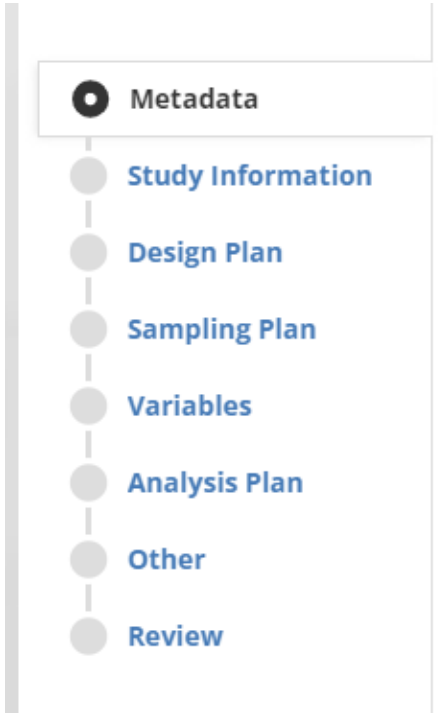
OSF Workflow

Create new registration



OSF Workflow: **Pre-**registration

- General information
- Hypothesis
- Design plan
  - Study type
  - Blinding
  - Study design
  - Randomization
- Sampling plan
  - Existing data*
  - Explanation of existing data
  - Data collection procedure*
  - Sample size* + rationale + stopping rule
- Variables
  - Dependent variables
  - Independent variables*
  - Indexing?
- Analysis plan
  - Statistical models*
  - Transformations
  - Inference criteria (e.g., p-value)
  - Data exclusion
  - Missing data
  - Exploratory analysis



Continue your registration by selecting a registration form:

- OSF Preregistration
- Open-Ended Registration
- Qualitative Preregistration
- Secondary Data Preregistration
- Generalized Systematic Review Registration
- Registered Report Protocol Preregistration
- OSF-Standard Pre-Data Collection Registration
- Preregistration Template from AsPredicted.org
- Replication Recipe (Brandt et al., 2013): Post-Completion
- Replication Recipe (Brandt et al., 2014): Pre-Registration
- Pre-Registration in Social Psychology (van 't Veer & Giner-Sorolla, 2016): Pre-Registration

- Registration prior to creation of data
- Registration prior to any human observation of the data
- Registration prior to accessing the data
- Registration prior to analysis of the data
- Registration following analysis of the data

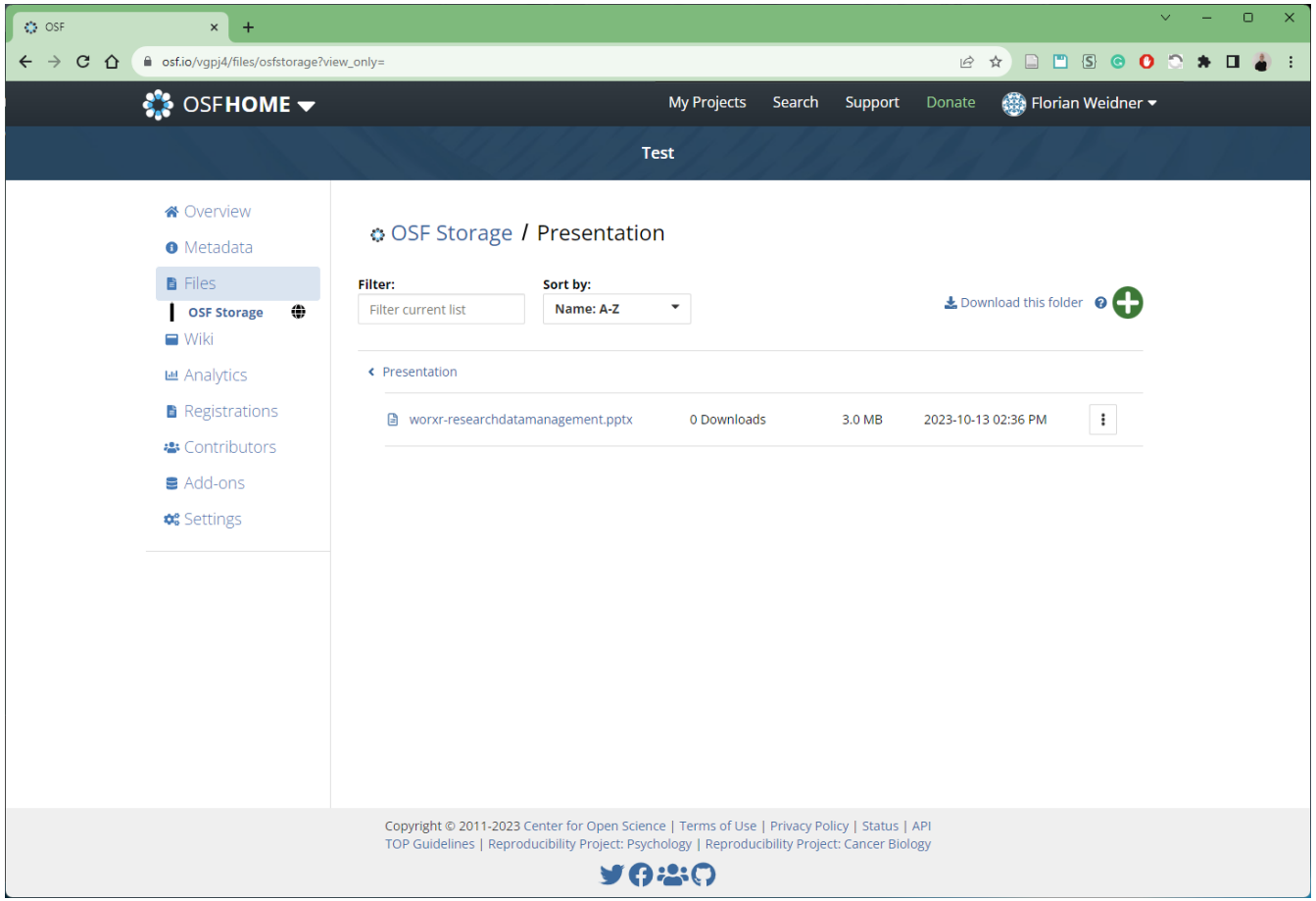
Study design *

Describe your study design. The key is to be as detailed as is necessary given the specific parameters of the design. There may be some overlap between this question and the following questions. That is OK, as long as sufficient detail is given in one of the areas to provide all of the requested information. Examples include two-group, factorial, randomized block, and repeated measures. Is it a between (unpaired), within-subject (paired), or mixed design? Describe any counterbalancing required.

… do study …

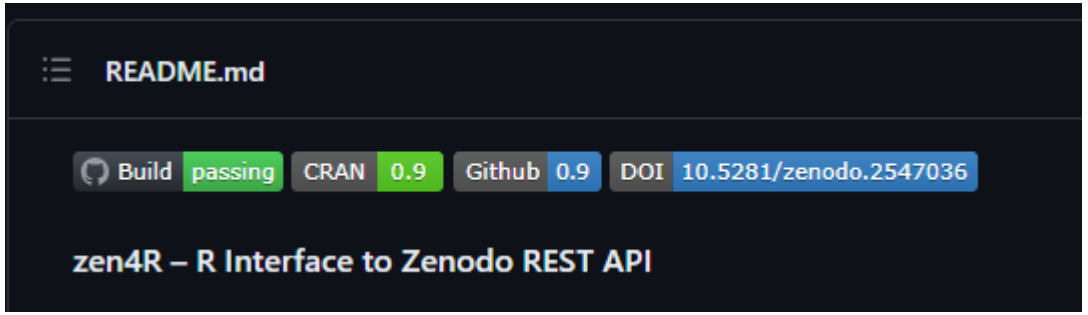
OSF Workflow

Add data.



OSF comments

- Free
- Foundation well-funded (long-term perspective)
- Preregistration
- Anonymous links
- Versioning possible
- 50GB
- Links to github release
  → no file mirror.
- Covers research lifecycle



https://github.com/eblondel/zen4R

Tl;dr: OSF

1. Create and fill out pre-registration.
2. Do study.
3. Upload data, analytics code, materials, papers, and supplemental.
4. Create view-only links (private or not)
5. Add them to manuscript.
6. Be happy ☺

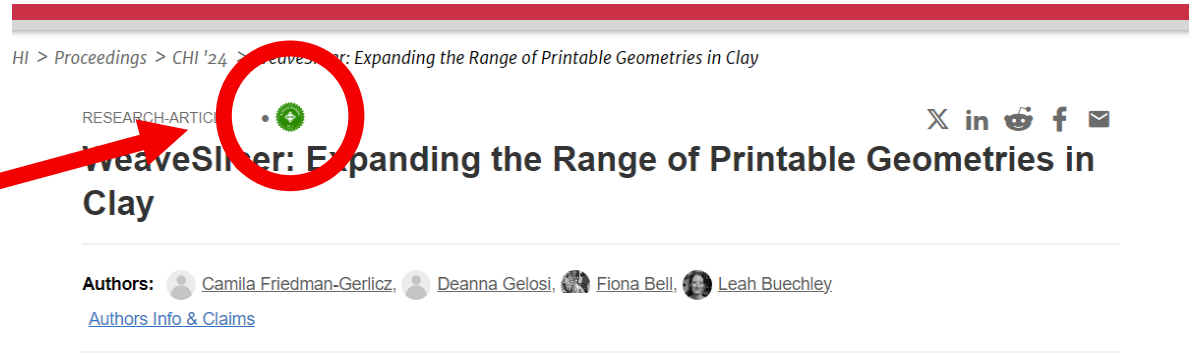
Example: IEEE VR 24 submission:
https://osf.io/uf46b/

ACM / PCS / TAPS

ACM: PCS / TAPS

Pro:

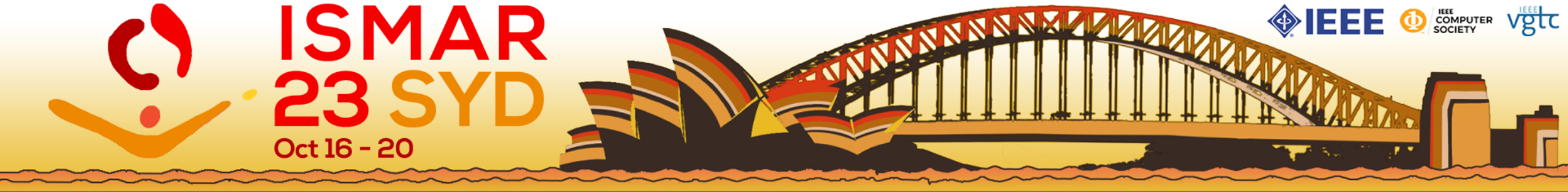
- You get a cool badge
- Discoverability



Con:

- File size limited:
  - PCS = 300 MB per file, 3 max
  - TAPS = 2GB max

Summary



All together?

- Soft requirement
  - Executable + setup instructions
  - Study protocol and self-defined questions
  - Raw data + scripts to calculate results in paper
- OSF: allrounder, free, pre-registration, 50GB
- Zenodo: Github, free, 50GB, no pre-registration
- ACM / PCS / TAPS: Free, limited file size

What now?

- Make materials available.
- What we **should do**: OSF with pre-reg (time, tho)
- What I **actually do**:
  - Submission:
    - Upload most important stuff to PCS / TAPS
    - Reserve and link Zenodo/DOI in paper with comment (on publication)
  - After acceptance
    - All on Zenodo
    - Link/DOI in paper

Replicability and data/algorithm storage and availability

3rd Workshop on Replication in Extended Reality (WoRXR)

Florian Weidner, Lancaster University

Thank you! Questions?

ISMAR 23 SYD
Oct 16 - 20

IEEE

IEEE COMPUTER SOCIETY

IEEE vgtc

Some resources

- https://www.acm.org/publications/policies/digital-artifacts
- https://journals.ieeeauthorcenter.ieee.org/create-your-ieee-journal-article/research-reproducibility/

FAIR data

- https://en.wikipedia.org/wiki/FAIR_data
- https://www.go-fair.org/fair-principles/