

Replicability and data/algorithm storage and availability

3rd Workshop on Replication in Extended Reality (WoR XR)

Florian Weidner

Lancaster University 

 IEEE  IEEE COMPUTER SOCIETY  vgtc



ISMAR
23 SYD
Oct 16 - 20



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](https://ieeedataport.org)
- Some institutions have platforms
- Criteria:
 - Offers DOI
 - promises long-term storage
 - Stable funding
 - (Free)



IEEEDataPort[™]

zenodo

Zenodo

All about storage.



ISMAR
23 SYD
Oct 16 - 20



Zenodo UI

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

Digital Object Identifier*

Do you already have a DOI for this upload? Yes No

Copy/paste your existing DOI here...

A DOI allows your upload to be easily and unambiguously cited. Example: 10.1234/foo.bar

The screenshot displays the Zenodo upload interface. On the left, a vertical menu contains several expandable sections: Files, Basic information, Recommended information, Funding, Alternate identifiers, Related works, References, Publishing information, and Conference. The 'Files' section is currently expanded, showing a 'Storage available' indicator (0 out of 100 files, 0 bytes out of 50.00 GB) and a central area for file uploads with a 'Drag and drop files' instruction and an 'Upload files' button. On the right side, there are two main panels. The top panel, titled 'Draft', includes 'Save draft' and 'Preview' buttons, and a prominent green 'Publish' button. The bottom panel, titled 'Visibility', shows 'Files only' with 'Public' selected over 'Restricted'. A green lock icon and the text 'Public The record and files are publicly accessible.' are displayed. Below this, an 'Options' section contains a checkbox for 'Apply an embargo' with a note: 'Record or files protection must be restricted to apply an embargo.'

Example

Published February 25, 2021 | Version 1.0.0 Software Open

S3D Dashboard

Weidner, Florian¹ 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](https://doi.org/10.5281/zenodo.4562267). This DOI represents all versions, and will always resolve to the latest one. [Read more](#).

External resources

Indexed in

OpenAIRE

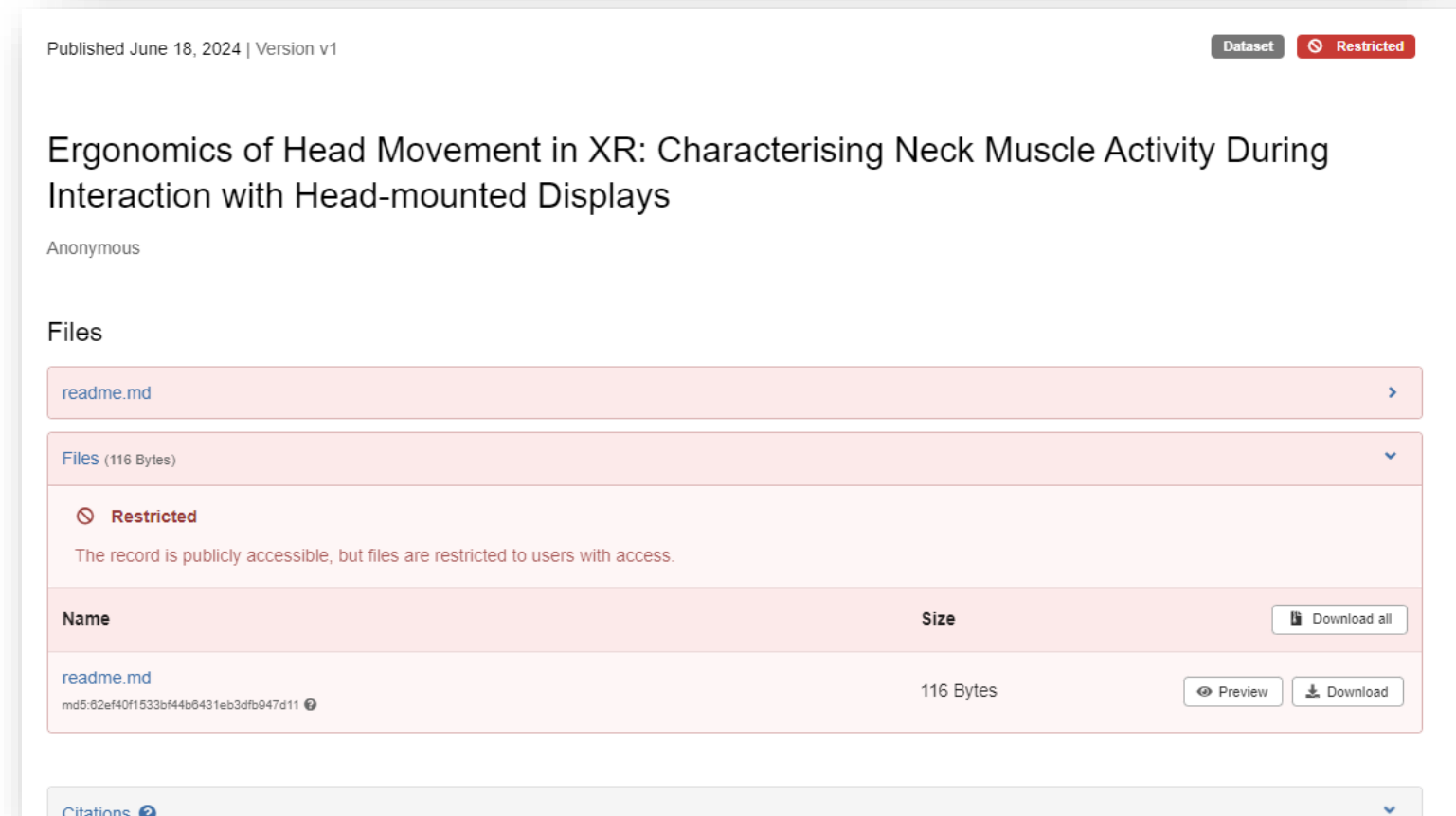
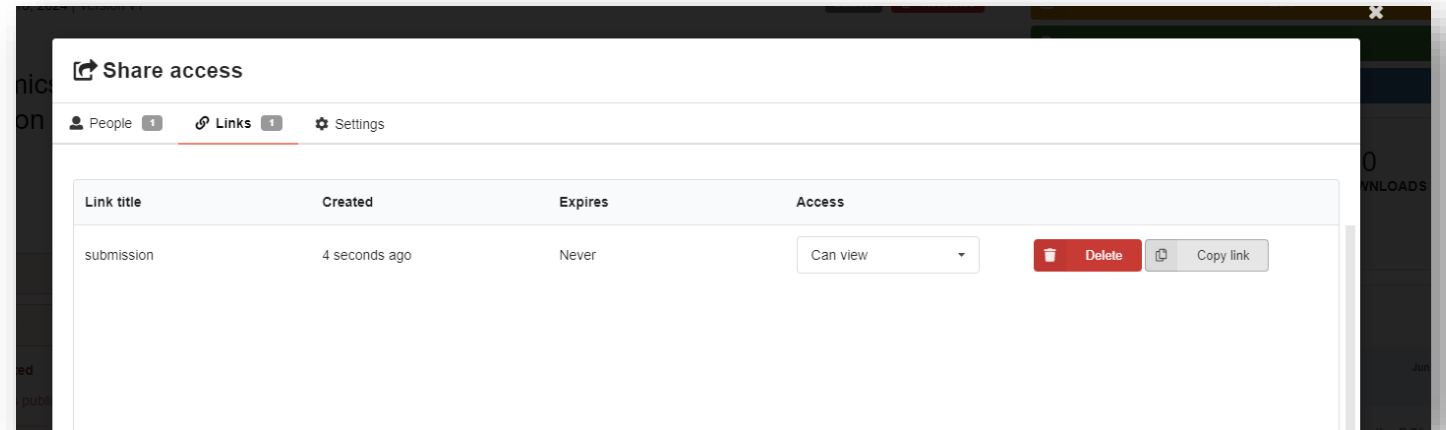
Communities

⌵

This record is not included in any communities yet.

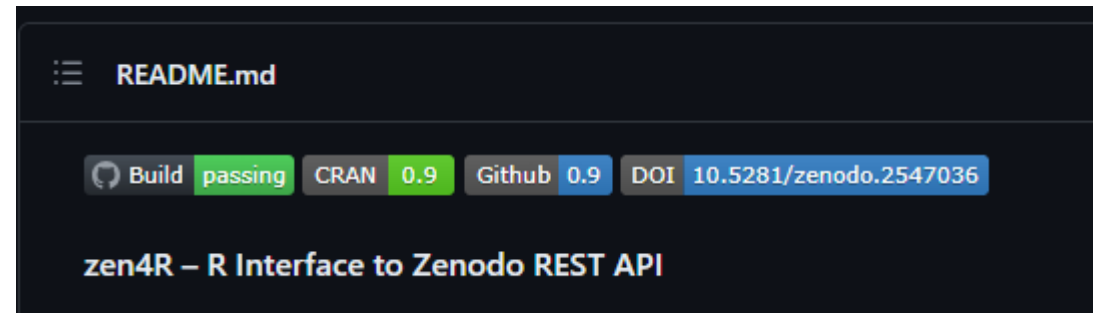
Private?

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)"



ISMAR
23 SYD
Oct 16 - 20

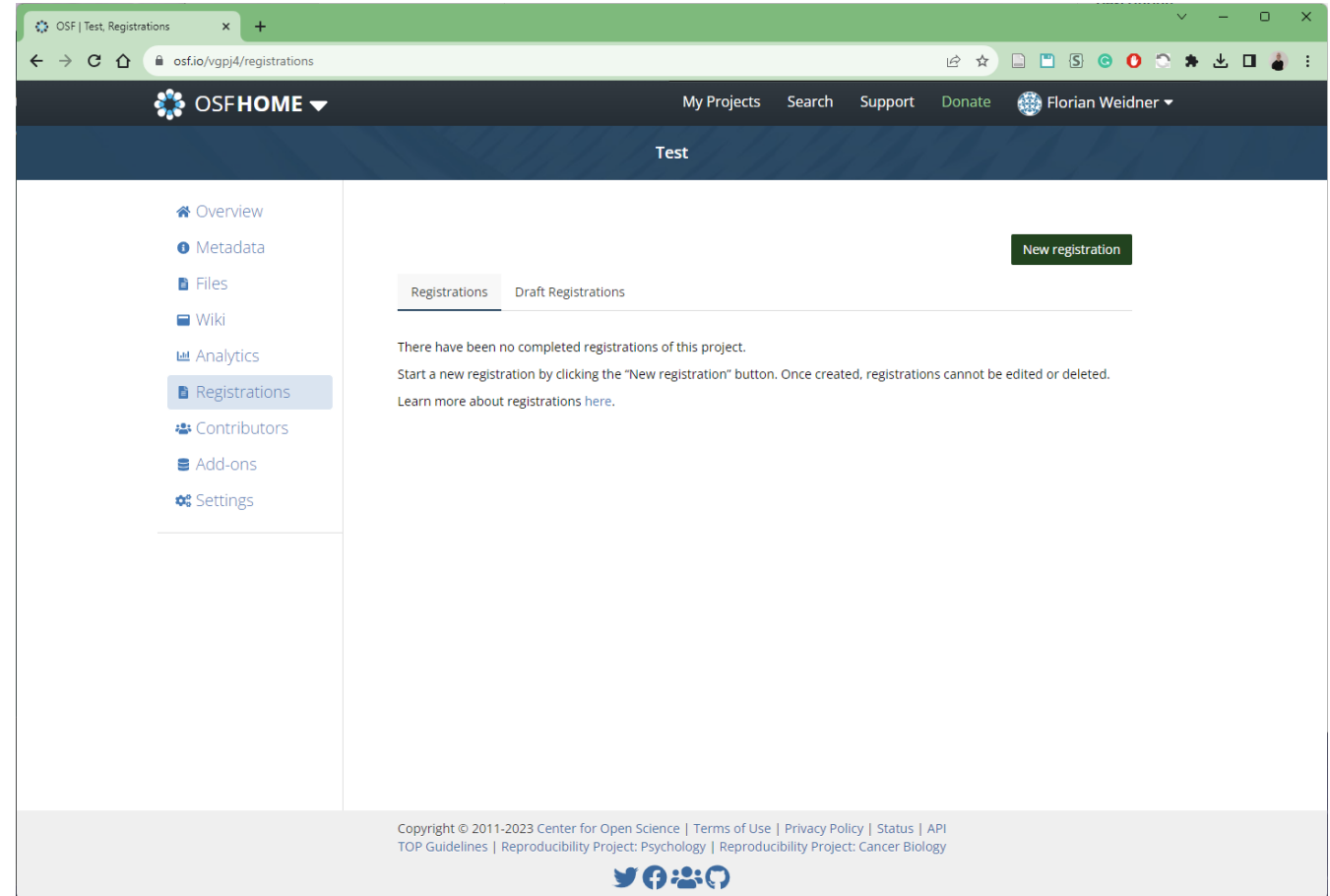


Registration + Project (+ Preprint)

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

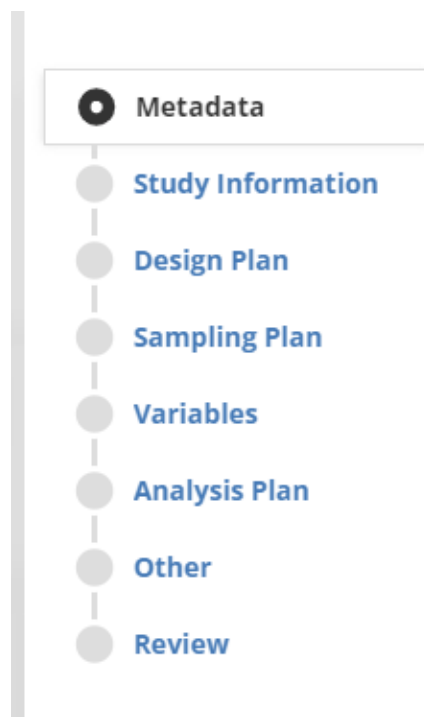
OSF Workflow

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 ?

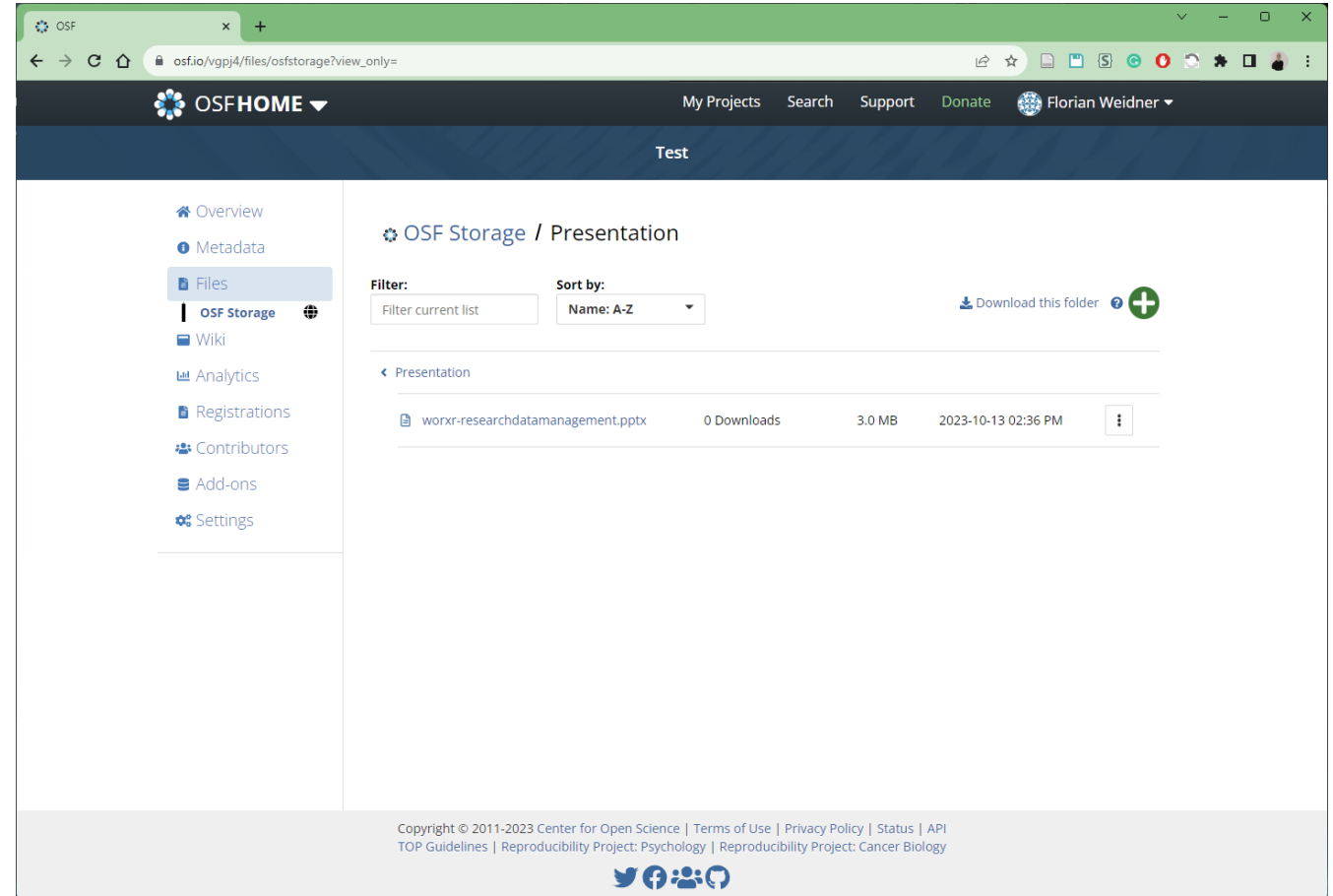
Study design *

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.



The screenshot shows the OSF Storage interface for a project named "Test". The browser address bar displays the URL `osf.io/vgpj4/files/osfstorage?view_only=`. The page header includes the OSF logo, "OSFHOME", and navigation links for "My Projects", "Search", "Support", "Donate", and the user profile "Florian Weidner".

The main content area is titled "OSF Storage / Presentation". It features a "Filter:" input field with the text "Filter current list" and a "Sort by:" dropdown menu set to "Name: A-Z". A "Download this folder" button with a plus icon is also visible.

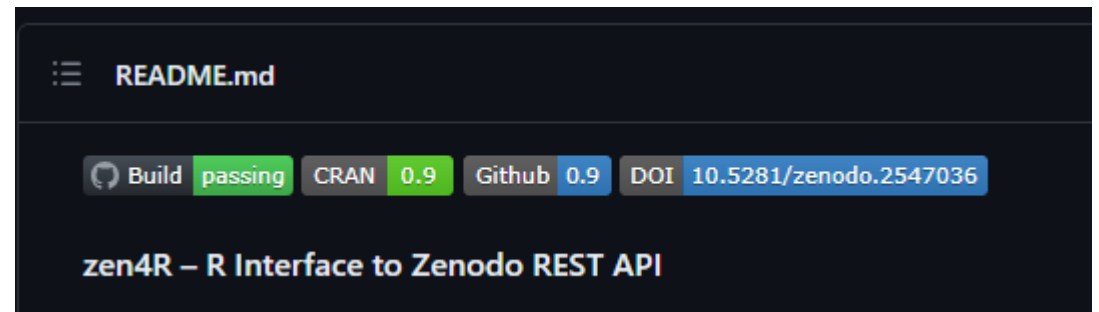
Below the filters, a table lists the contents of the "Presentation" folder:

Presentation			
 worxr-researchdatamanagement.pptx	0 Downloads	3.0 MB	2023-10-13 02:36 PM

The footer contains copyright information: "Copyright © 2011-2023 Center for Open Science | Terms of Use | Privacy Policy | Status | API TOP Guidelines | Reproducibility Project: Psychology | Reproducibility Project: Cancer Biology" and social media icons for Twitter, Facebook, and LinkedIn.

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



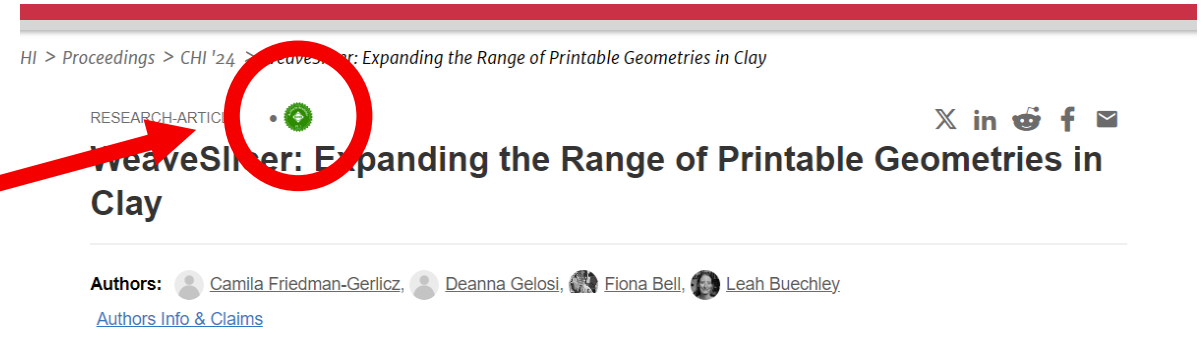
ISMAR
23 SYD
Oct 16 - 20



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



ISMAR
23 SYD
Oct 16 - 20



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 (WoR XR)

Florian Weidner, Lancaster University

Thank you! Questions?



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/>

F A I R



Findable



Accessible



Interoperable



Reusable