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)

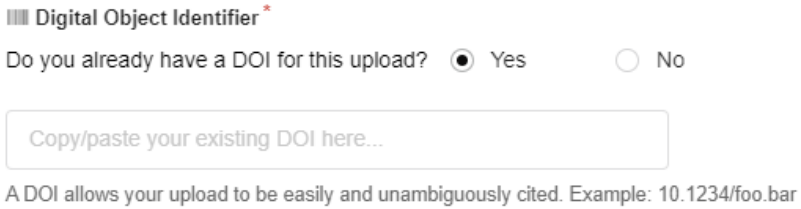
Zenodo

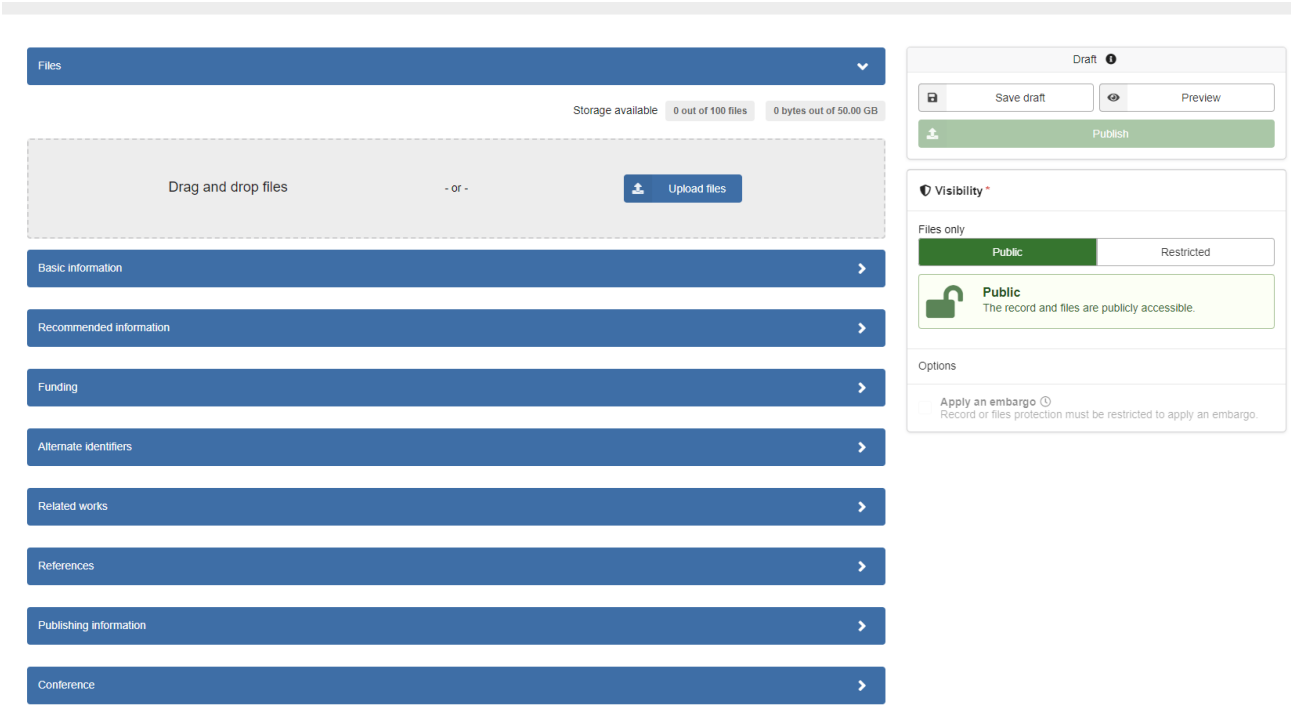
All about storage.

Zenodo UI

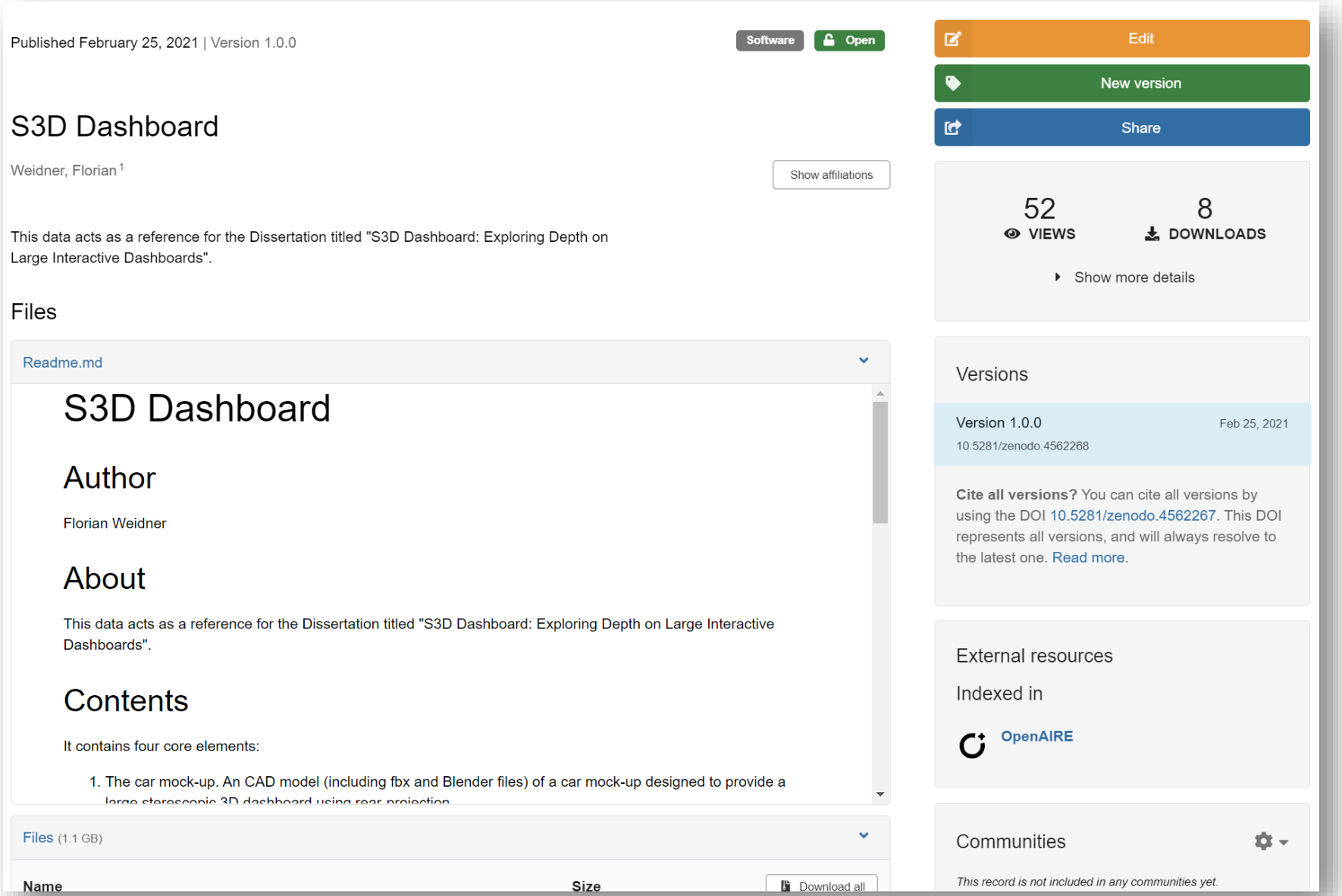
- Signup / login
- Create new upload
- Get DOI
- Fill in metadata
- Upload files
- Review
- Publish

➔ 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

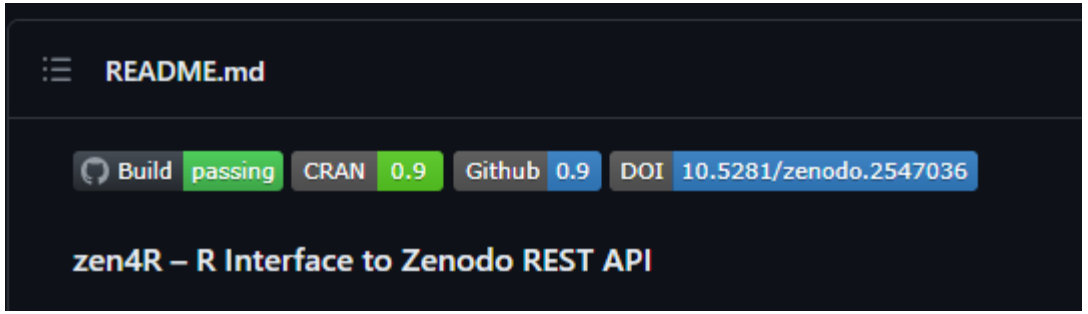
OpenAIRE

Communities

This record is not included in any communities yet.

Comments on Zenodo

- Free
- Funded by CERN (long-term perspective)
- No 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

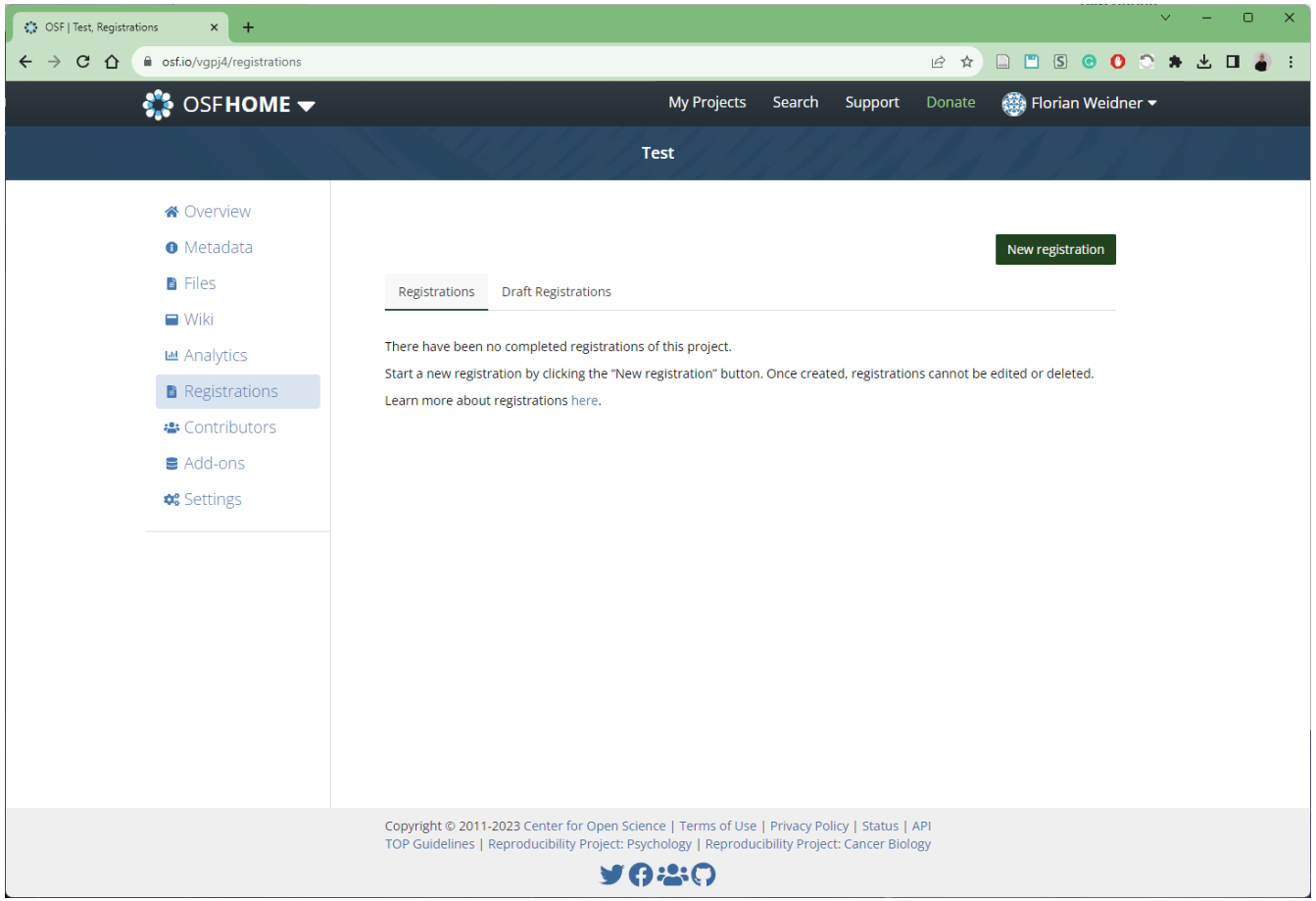
TIL: “OSF is maintained and developed by the Center For Open Science (COS)”

Registration + Project (+ Preprint)

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

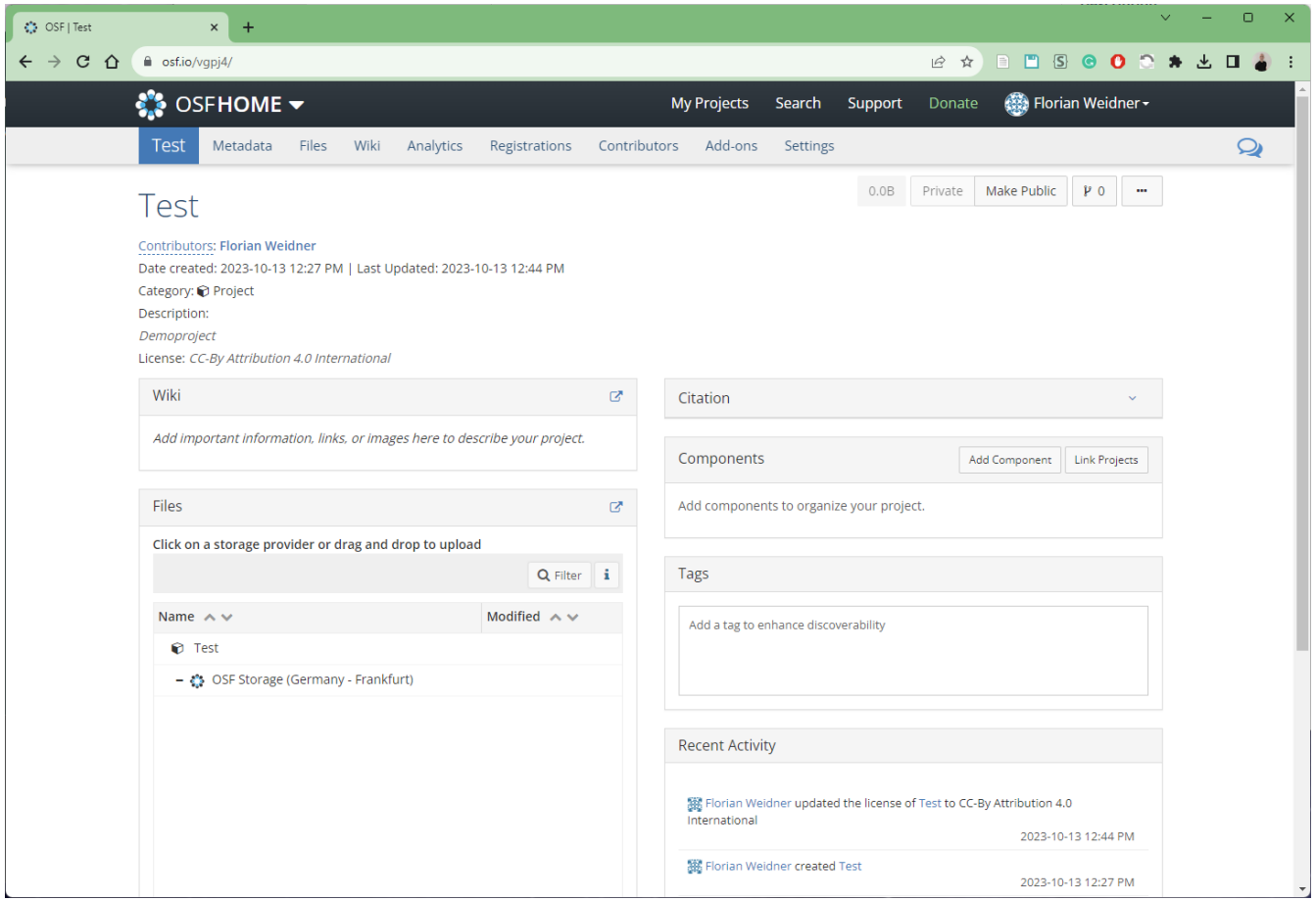
OSF Workflow

Create new registration



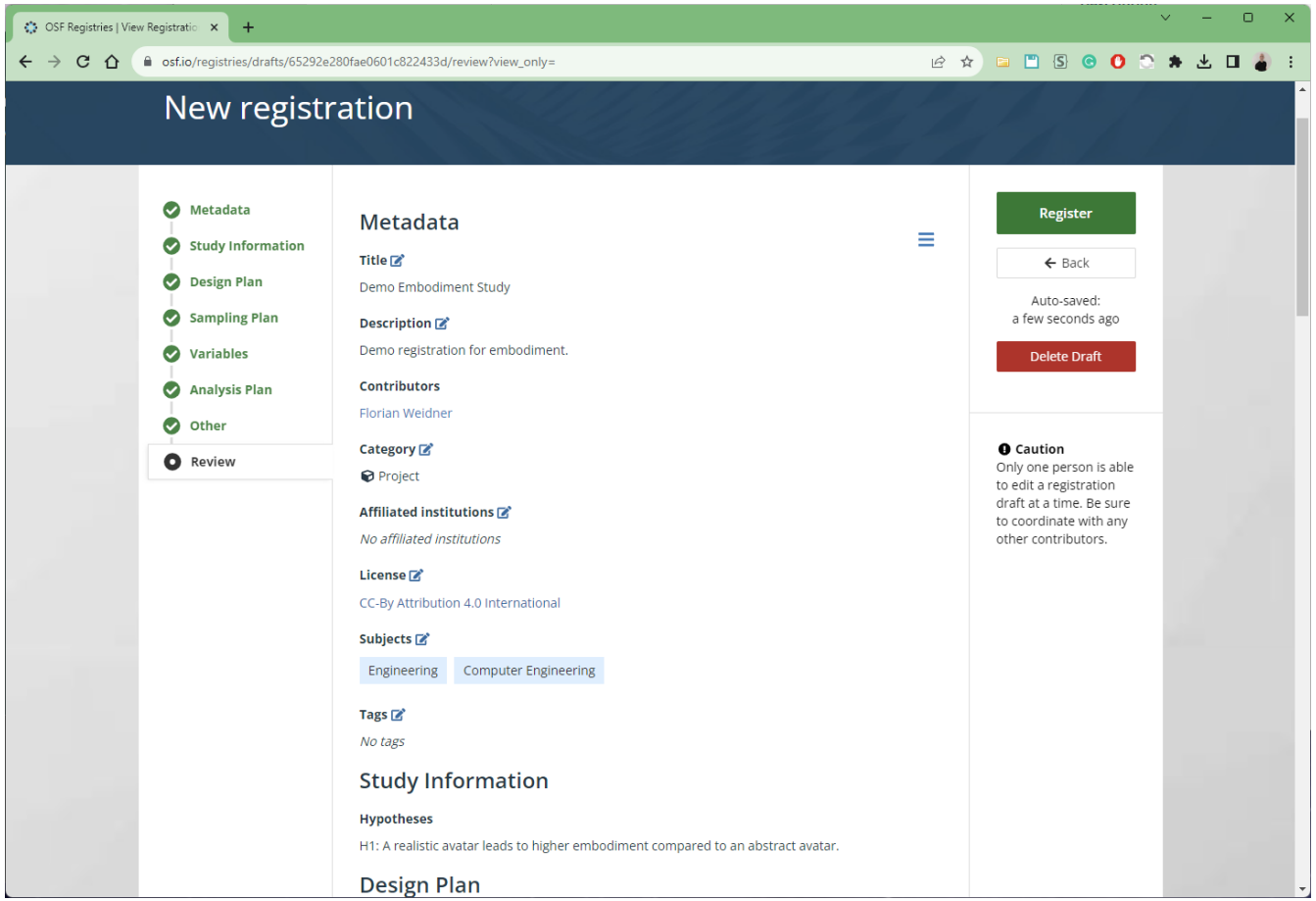
OSF Workflow

Creates new project



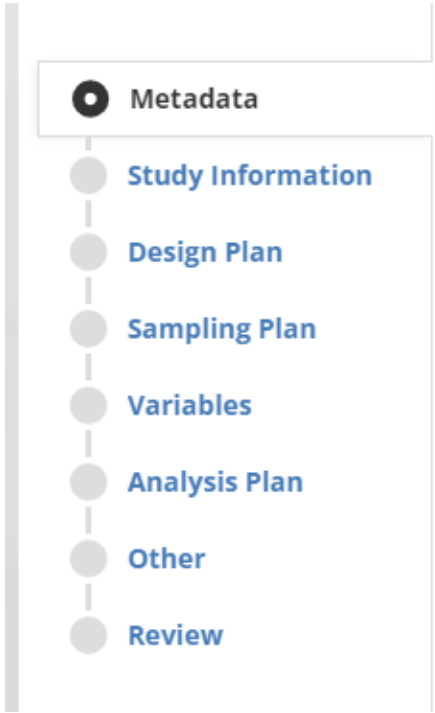
OSF Workflow

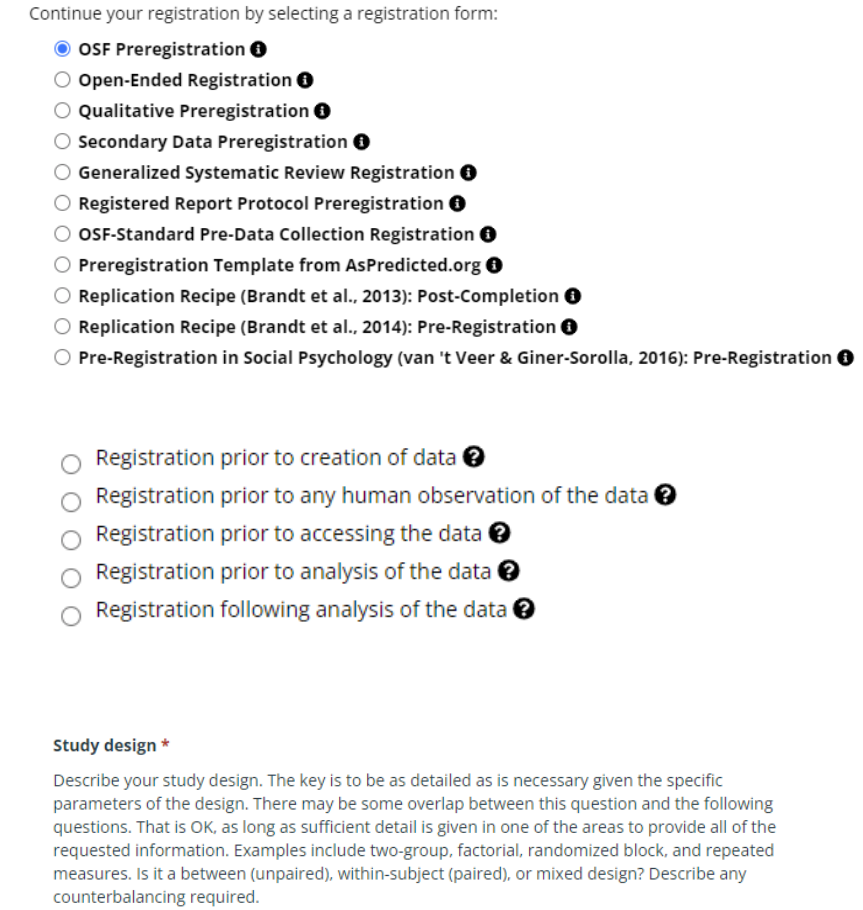
Fill out all info and register (updatable)



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

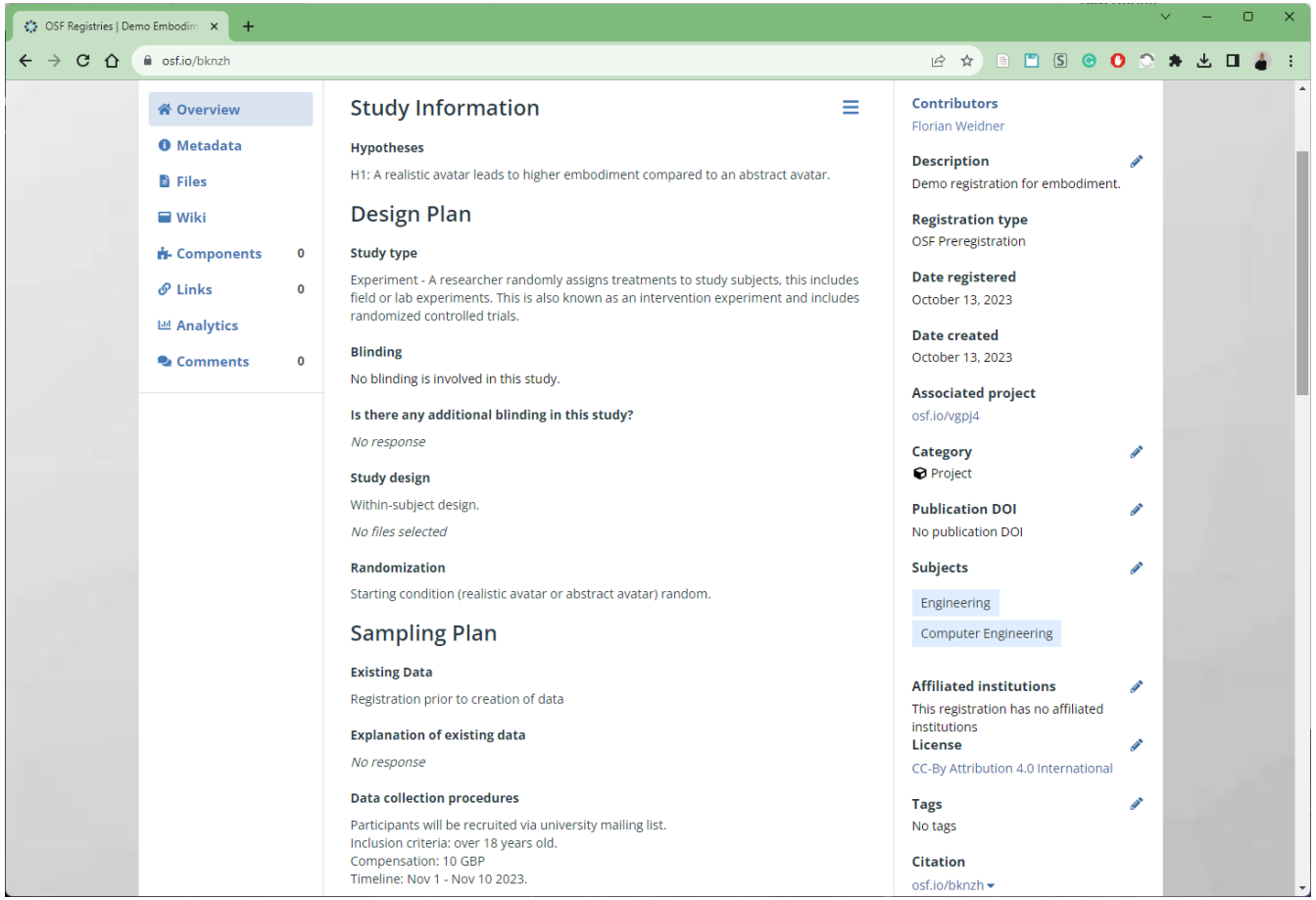




OSF Workflow

Anonymous link to registration possible: https://osf.io/bknzh/?view_only=bda587ab846f416482460924e019bf91

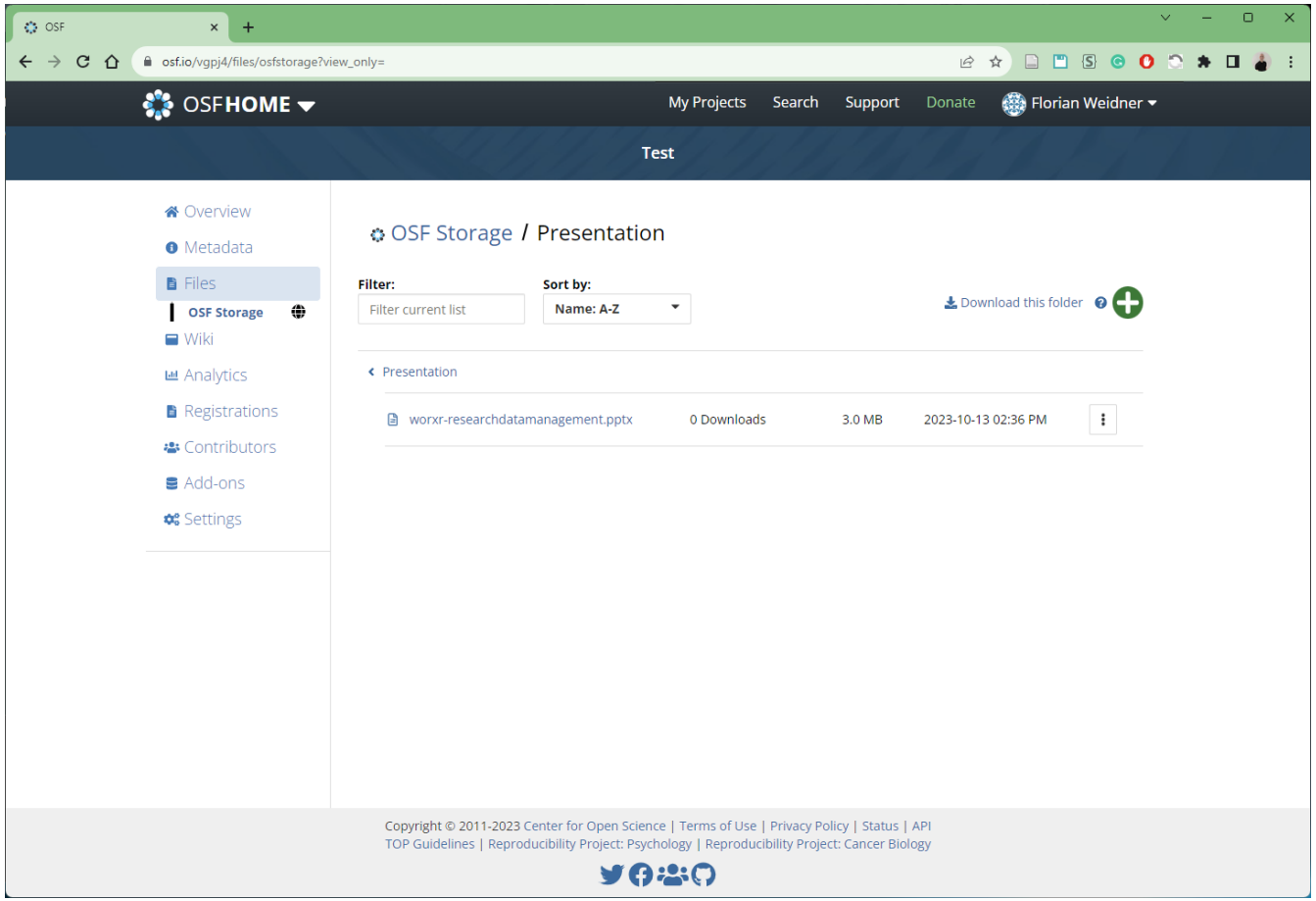
https://tinyurl.com/worxr1



… do study …

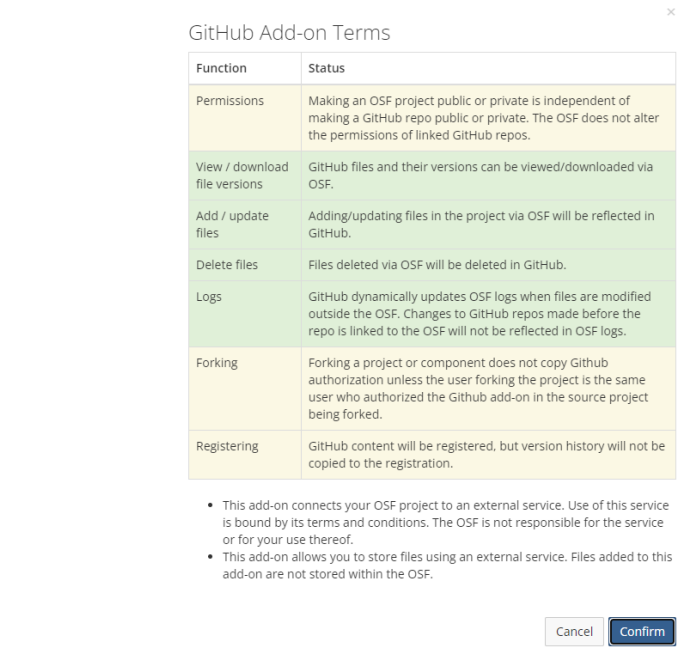
OSF Workflow

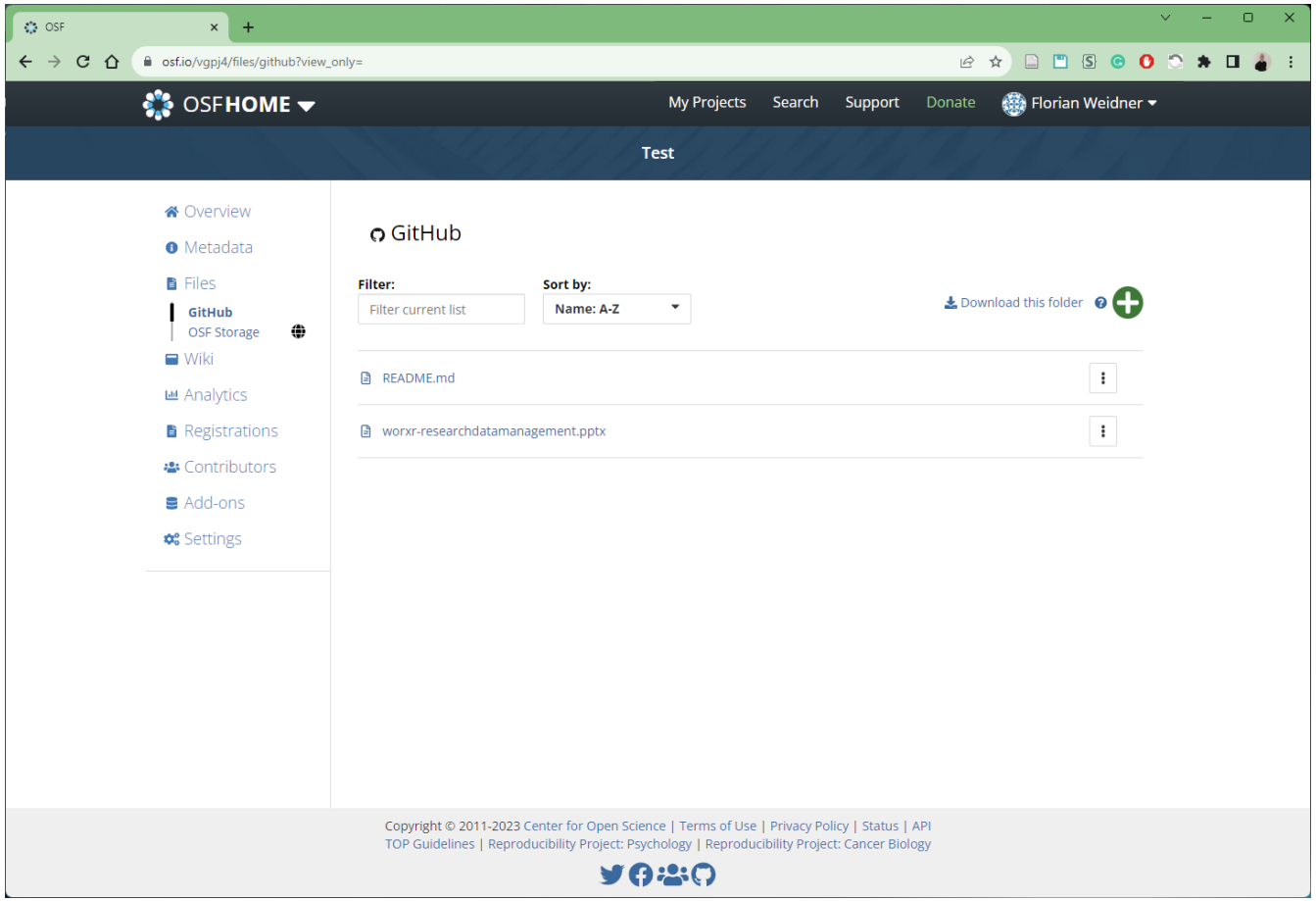
Add some data.



OSF Workflow

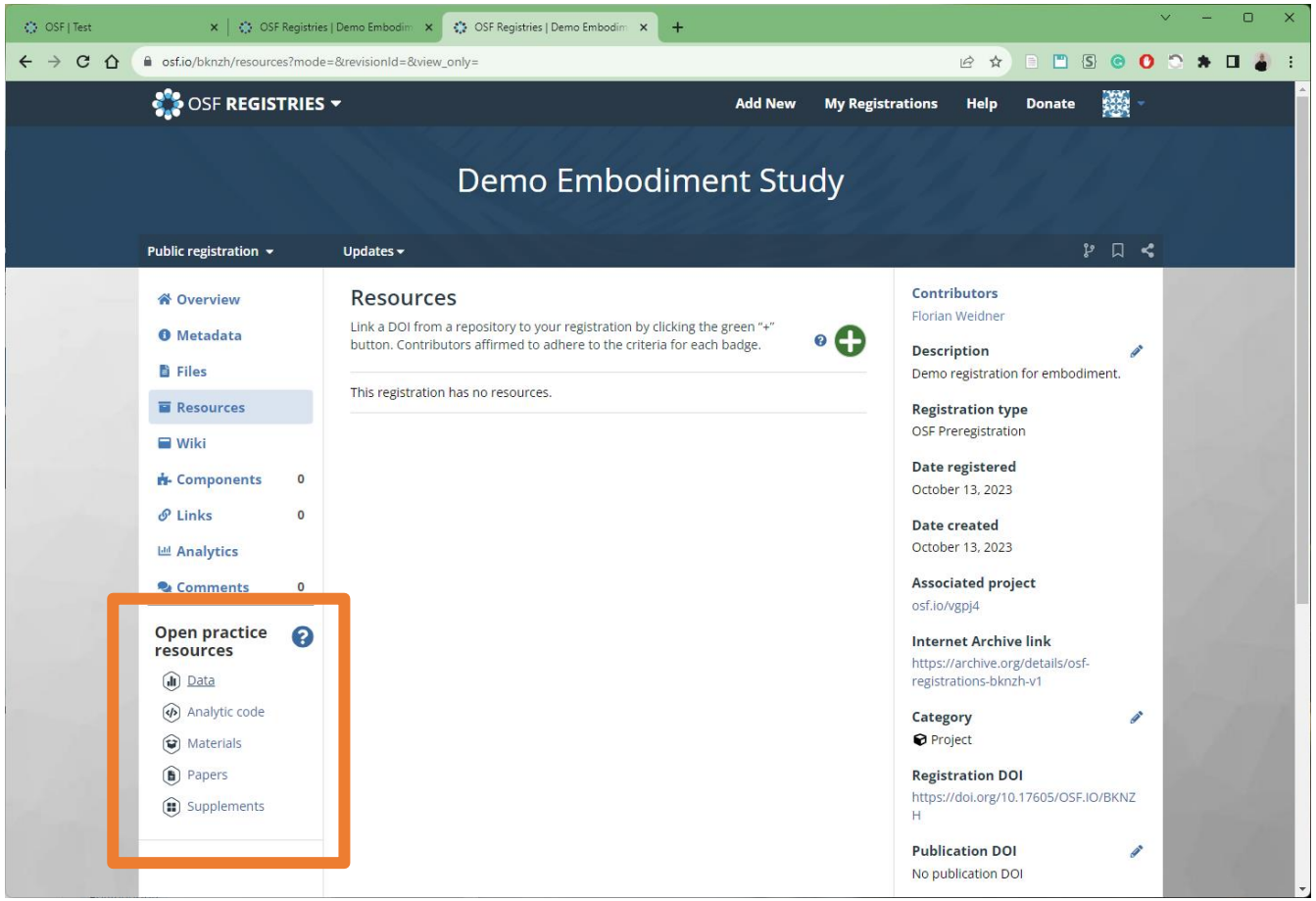
- Github + OSF
- No file mirror!





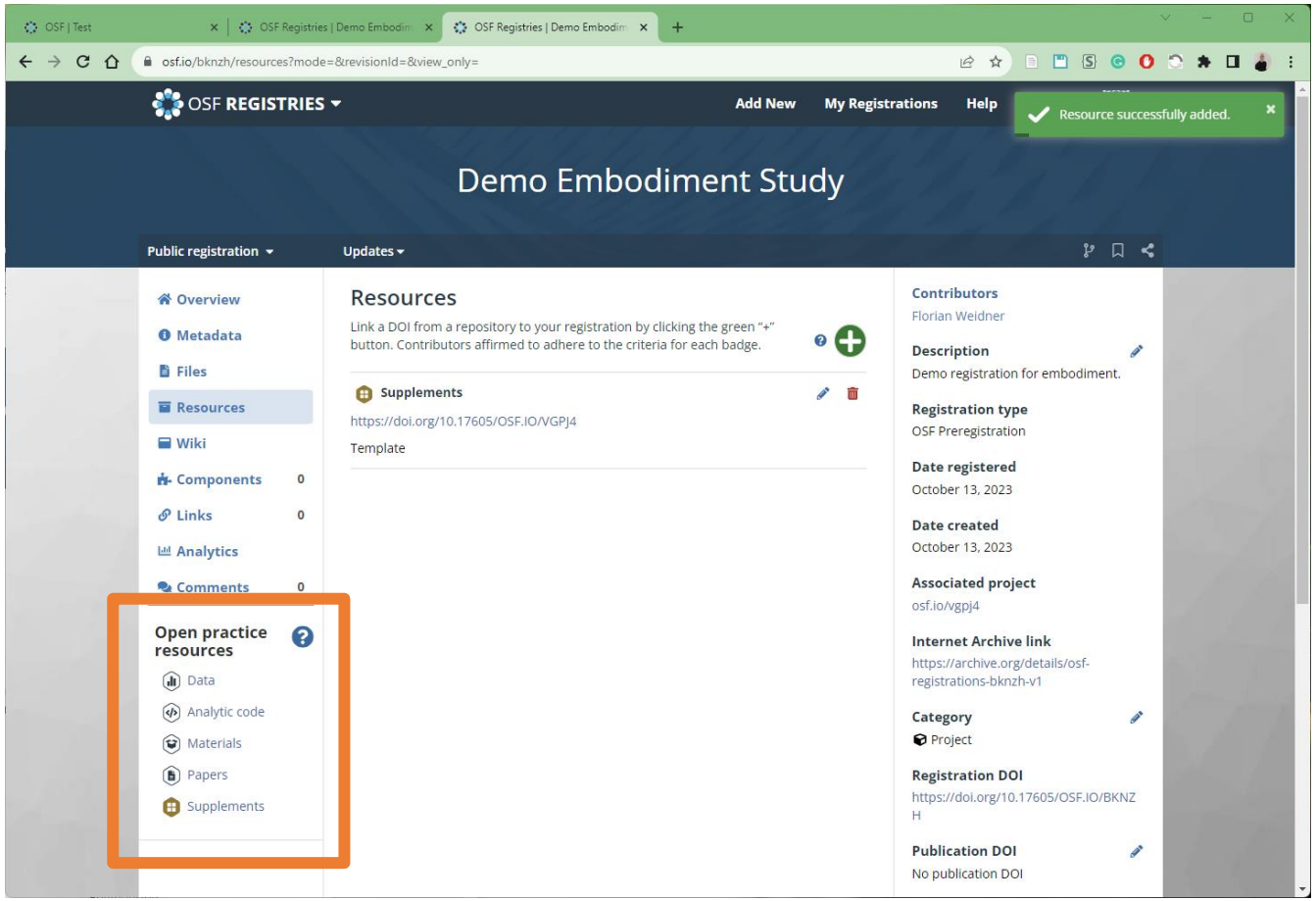
OSF Workflow

Link project to registration via project DOI.



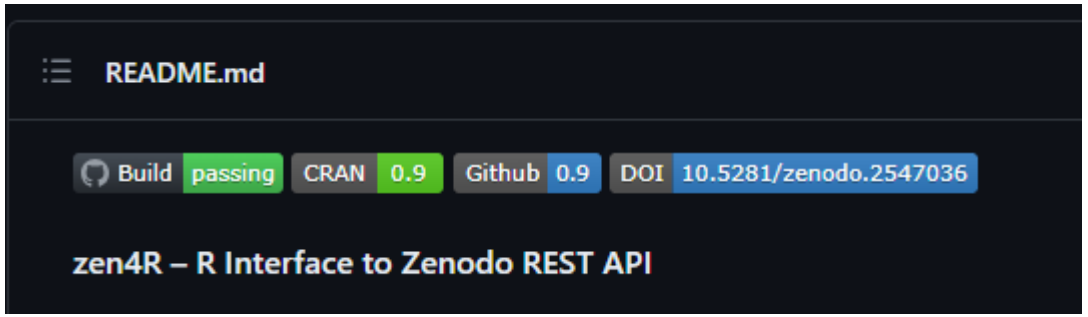
OSF Workflow

- Link project to registration via project DOI ☺
- [1] F. Weidner, “Demo Embodiment Study”, 13-Oct-2023. [Online]. Available: https://osf.io/bknzh.



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



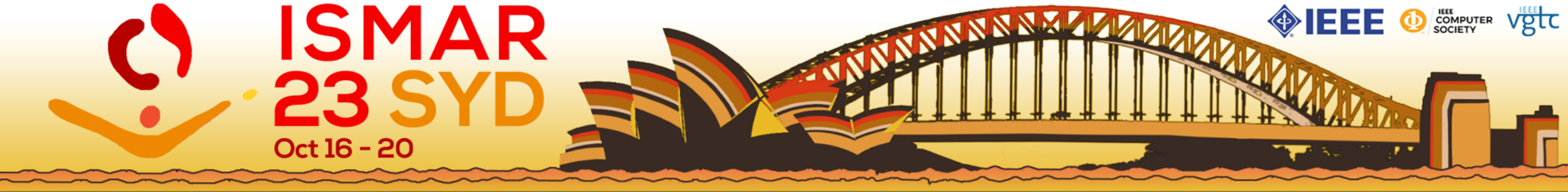
https://github.com/eblondel/zen4R

Tl;dr: OSF

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

IEEE DataPort

All about datasets.

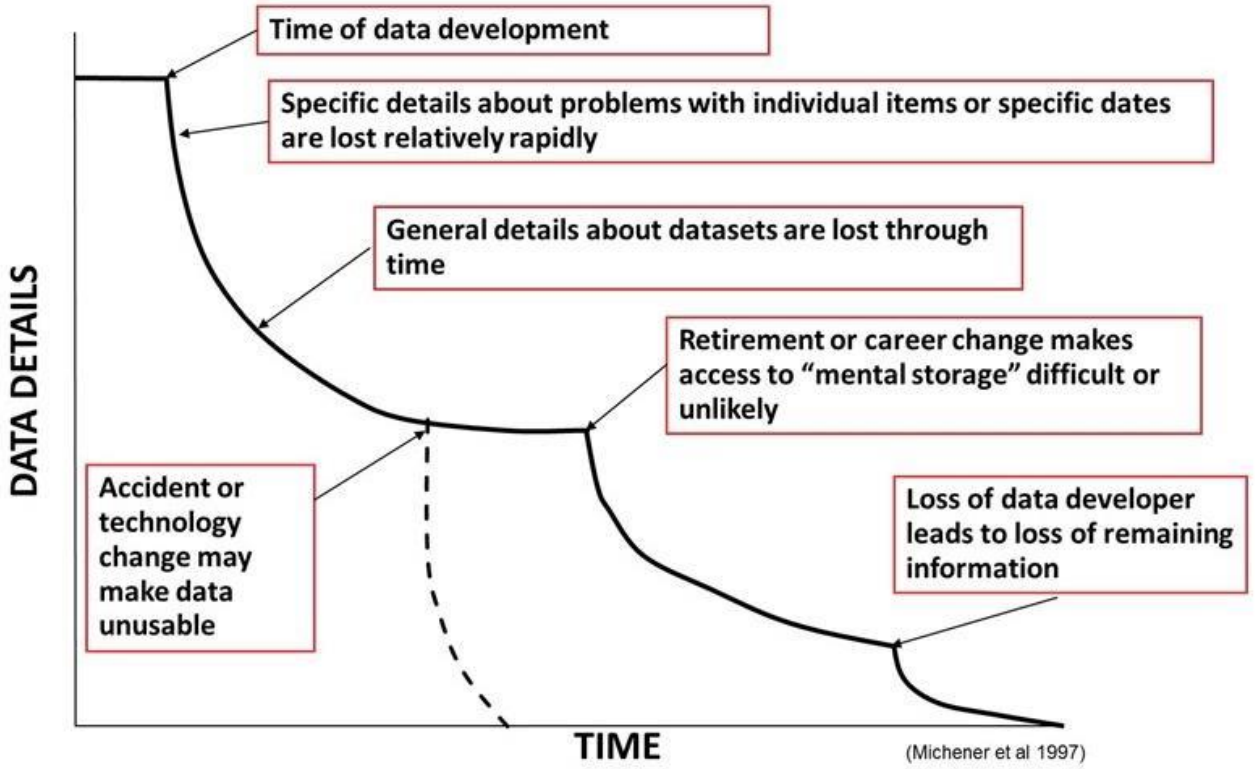


IEEE DataPort comments

- Free only for subscribers
- Up to 2TB or 10TB
- Open access for 1.950 USD (sometimes discounts)
- Long-term via IEEE
- Linked to IEEE account
- No preregistration
- DOI
- (designed for competitions)

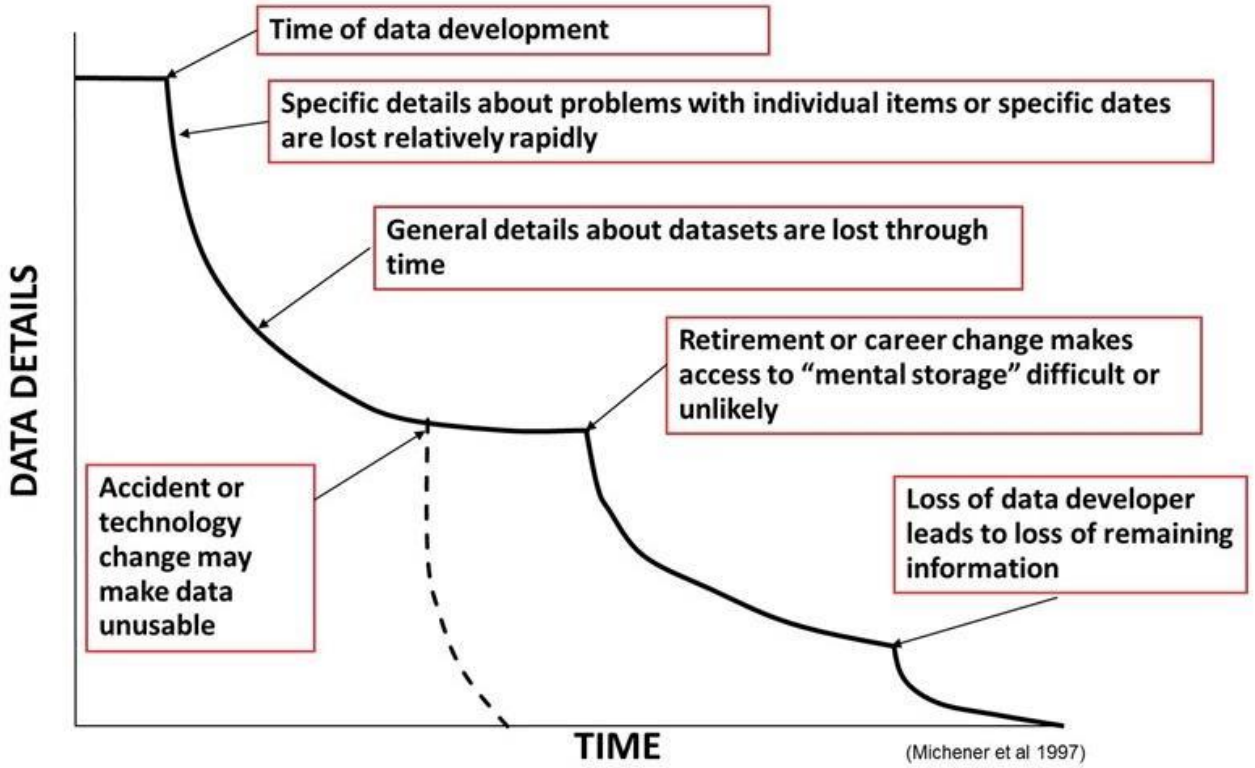
Data documentation

How do we document it?

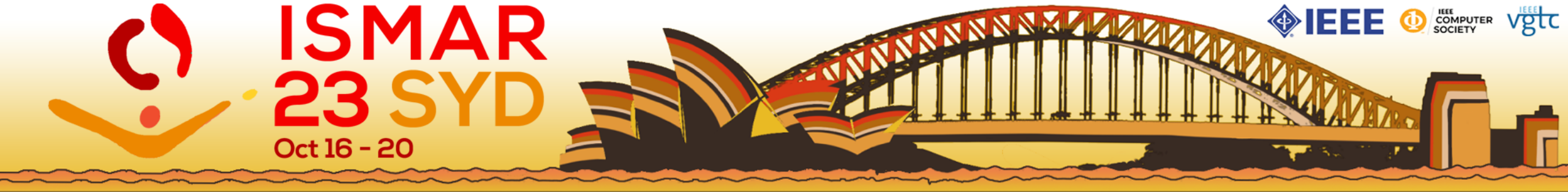


How do we document it?

- **README files**
- Version control
- Codebook
- Jupyter Notebooks
- Documented code
- Workflow screenshots
- …



Summary



All together?

- Soft requirement / extra recognition?
  - Executable + setup instructions
  - Study protocol and self-defined questions
  - Raw data + scripts to calculate results in paper
  - Video
  - Preprint
- OSF, Zenodo, IEEE DataPort?
  - OSF: allrounder, free, pre-registration, 50GB
  - Zenodo: easy via github, free, 50GB, no pre-registration
  - DataPort: large storage, paid

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/
- Fair != open