

# Open Science & Data Management

Harshvardhan Pandit  
ADAPT Centre, Dublin City University

[harshvardhan.pandit@adaptcentre.ie](mailto:harshvardhan.pandit@adaptcentre.ie)

<https://harshp.com/>



01 – Introduction

02 – Science

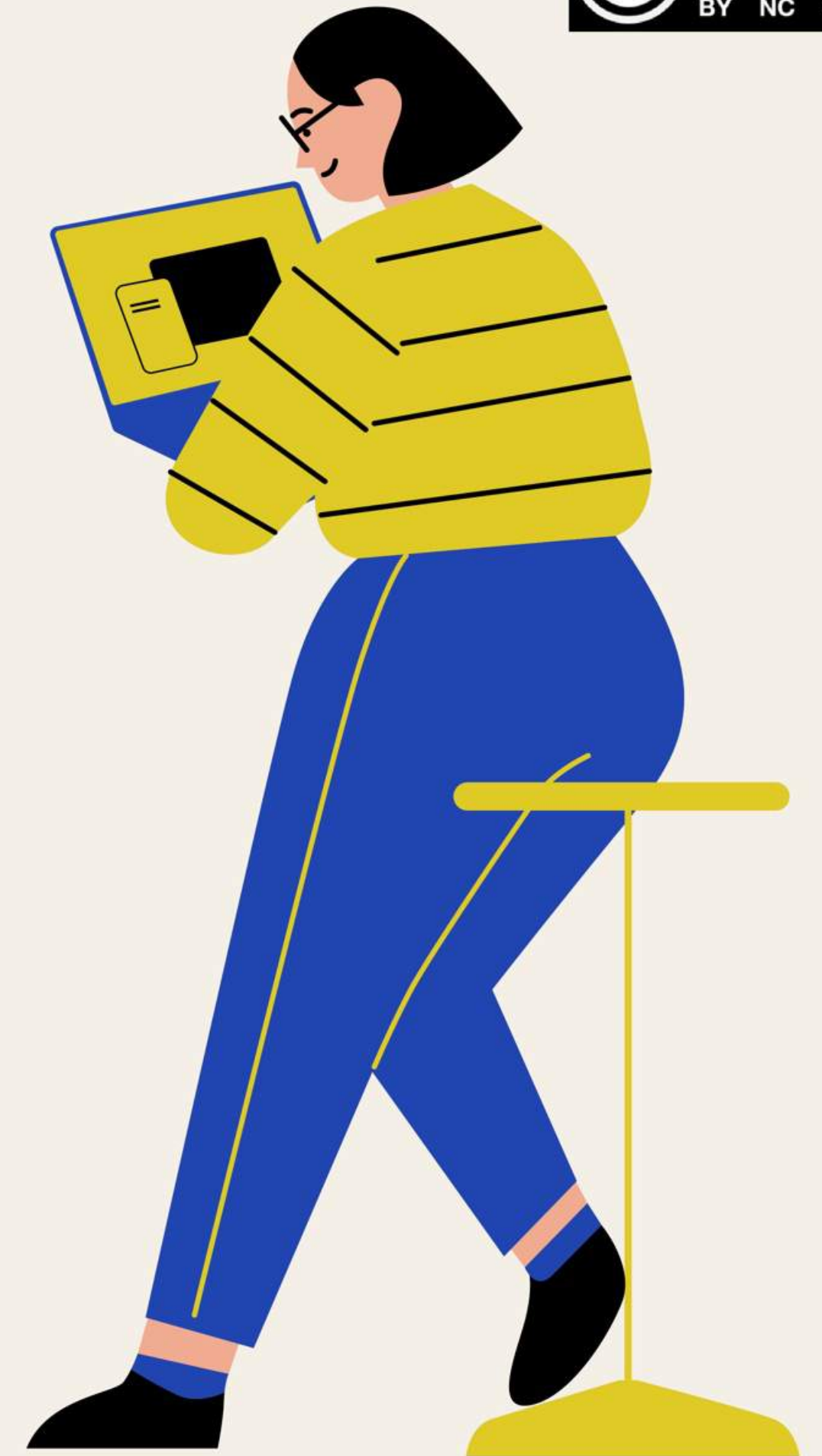
03 – Data

04 – Publishing

05 – Community

Open

Science / Data





# warm up



## Introductions



# Twitter

**@harsh@eupolicy.social**

@coolharsh55

moved to Mastodon as @harsh@eupolicy.social

📍 Dublin City, Ireland [🔗 harshp.com](https://harshp.com) 📅 Joined July 2009

207 Following 555 Followers



Edit profile

**harsh**

@harsh@eupolicy.social

Assistant Professor @ Dublin City University ; Chair W3C Data Privacy Vocabularies & Controls Community Group (DPVCG) ; Semantics x Privacy/DataProtection x Consent x GDPR

# Mastodon

# Google Scholar

**Harshvardhan J. Pandit** ✎

ADAPT Centre, Dublin City University (go to homepage copies)

Verified email at adaptcentre.ie - [Homepage](#)

[privacy](#) [semantic web](#) [consent](#) [GDPR](#) [regulatory compliance](#)



**Harshvardhan Pandit**

Assistant Professor at Dublin City University

County Dublin, Ireland · [Contact info](#)

489 connections

Open to

Add profile section

More



Dublin City University



Trinity College Dublin

# LinkedIn



# 02 – Science

Open science is the movement to make **scientific research** (including publications, data, physical samples, and software) and its dissemination accessible to all levels of society, amateur or professional.

Data

Visualization

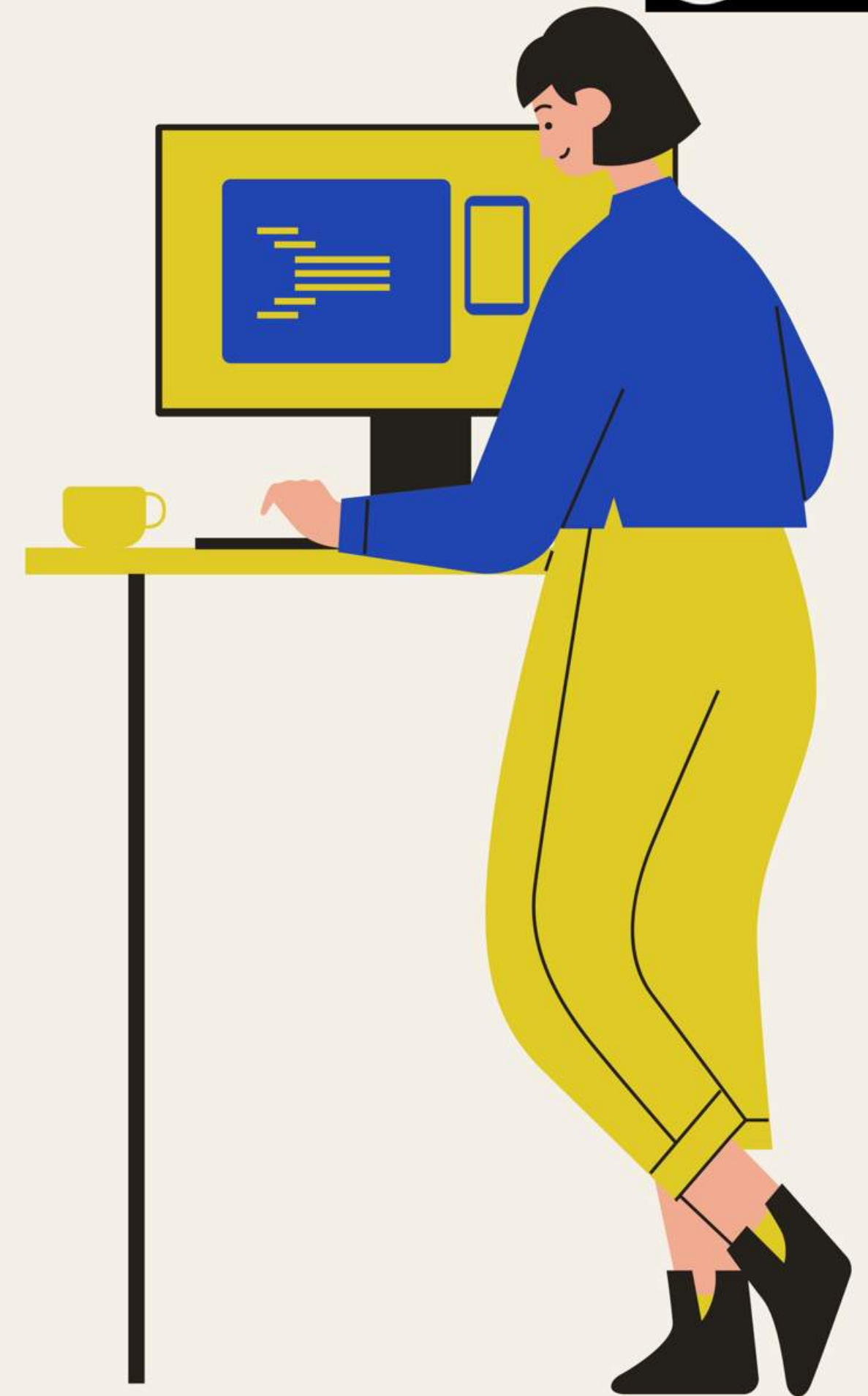
Open science is transparent and accessible knowledge that is shared and developed through collaborative networks. It encompasses practices such as publishing open research, campaigning for open access, encouraging scientists to practice open-notebook science (such as openly sharing data and code, broader dissemination and engagement in science and generally making it easier to publish, access and communicate scientific knowledge



# 02 – Six Principles of Open Science

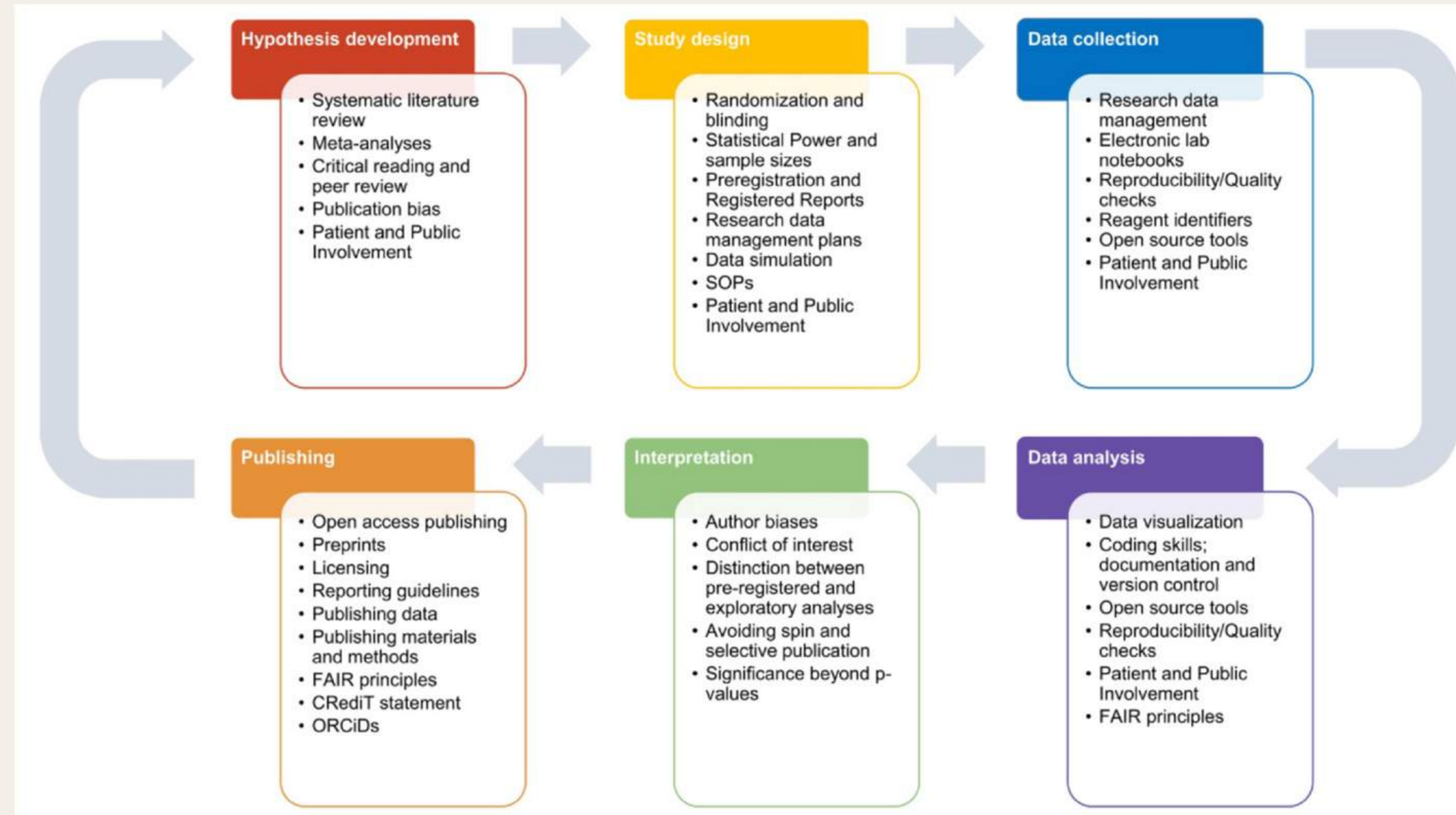
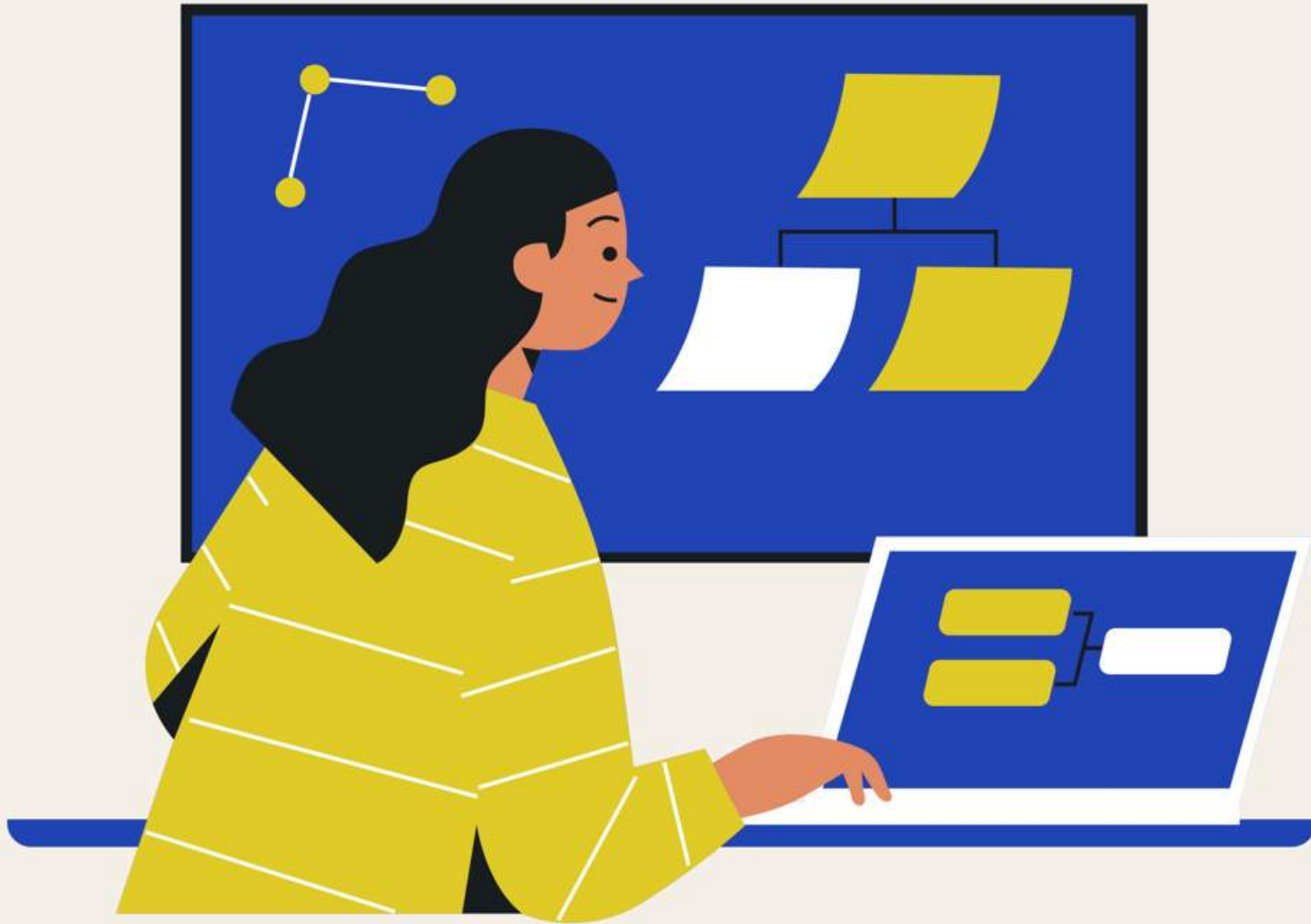
*Scientific process requires reviewing and replication*

- **Open methodology**
- **Open source**
- **Open data**
- **Open access**
- **Open peer review**
- **Open educational resources**





# 02 – Data



*Data or Information  
Samples or Collections  
Records or Documents*



# 03 – Data what now?

Data Spaces?



<https://ec.europa.eu/research/openscience/>



<http://www.orfg.org>



<http://scross.org>

## European Open Science Cloud (EOSC)

This is a cloud for research data in Europe. Background, policy information, events and publications related to the EOSC

## Open Science Policy Platform

Group that advises the Commission on how to develop open science policy. Meeting reports, member details and background

## Open science monitor

Tracking trends for open access, collaborative and transparent research across countries and disciplines.





# 04 - Publishing

arXiv [arxiv.org](https://arxiv.org)



[osf.io](https://osf.io)

[zenodo.org](https://zenodo.org)

zenodo

public “preprint server”

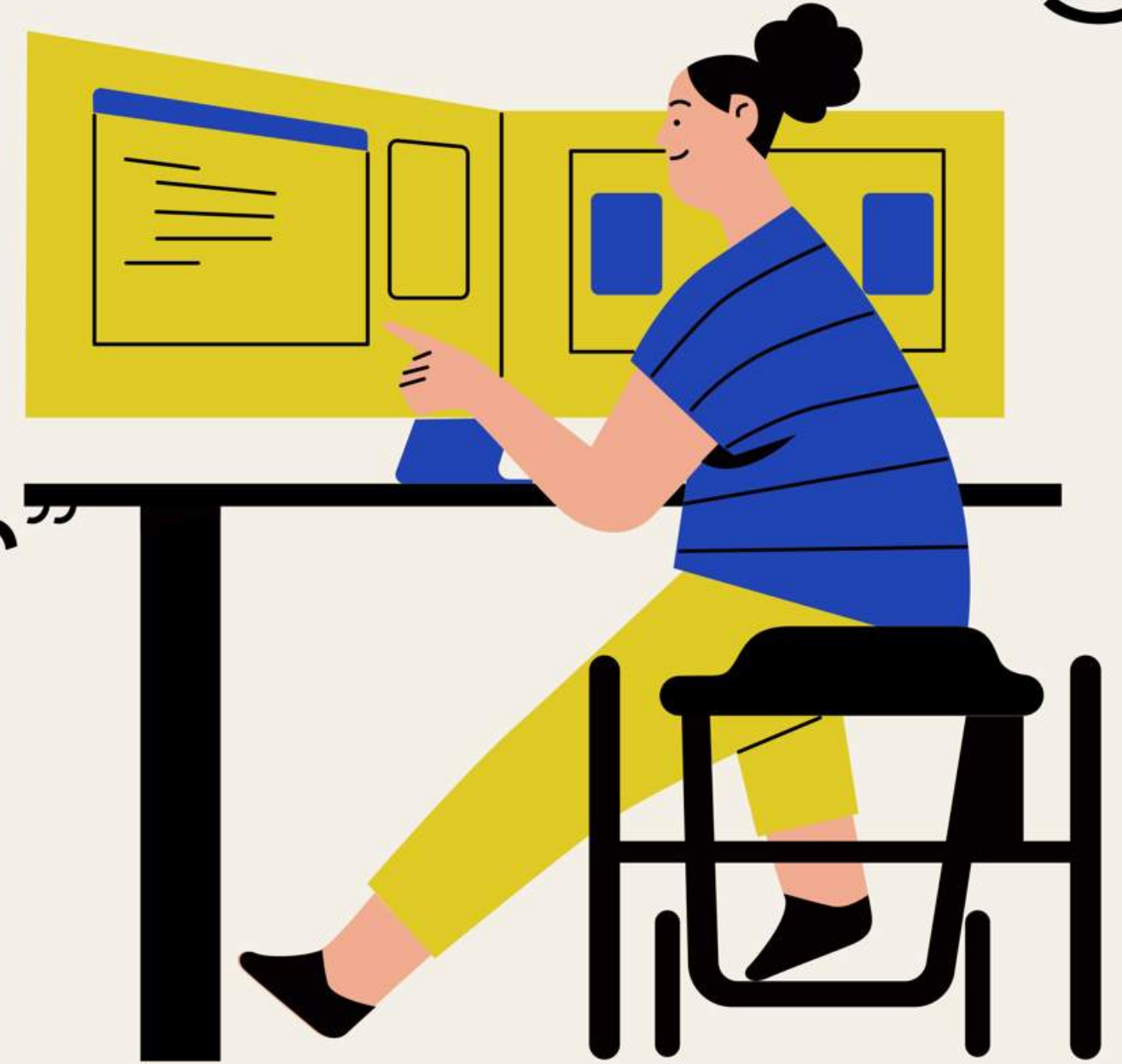


[tara.tcd.ie](https://tara.tcd.ie)



[doras.dcu.ie](https://doras.dcu.ie)

institutional repository



*Publications  
Reports  
Data*





# 04 – Publishing Options

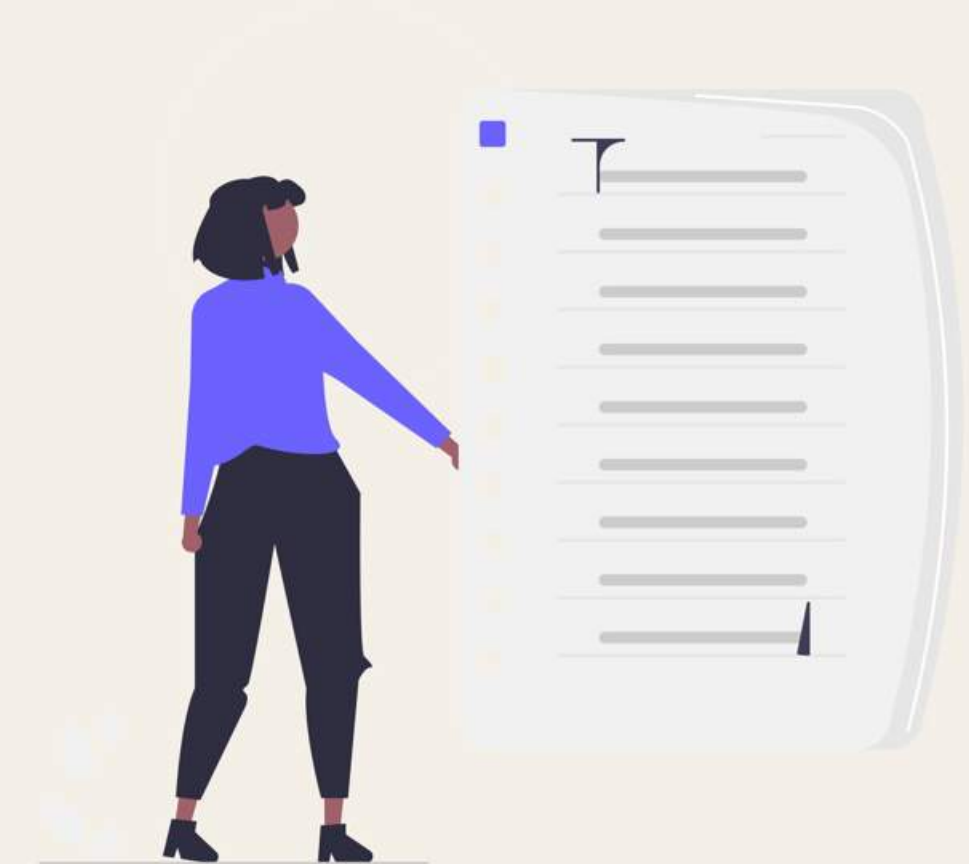
Draft --> Publish for comments  
Finalise --> Publish for feedback  
Accepted --> Publish for early access  
Presented --> Publish slides  
Published --> Publish for 'open' copies

*Publishing articles at different stages helps different “aims” in the context of Open Science*

*Can you cite “not accepted” articles ??? What about “accepted but not published” ???*

No

Yes





# O4 – FAIR Publishing

## FINDABLE

Unique identifiers and metadata are used to allow data to be located quickly and efficiently



## ACCESSIBLE

Data is open, free and universally available for research discovery efforts



## INTER-OPERABLE

A common programming language is used to allow use in a broad range of applications



## REUSABLE

All data is clearly described and outlines associated data-use standards





# O4 – Publishing Open Science

## Key Pillars of Open Science



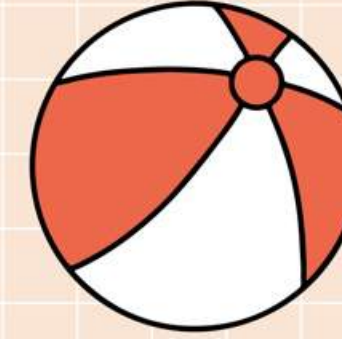
**Open Scientific Knowledge:** scientific publications, research data, software, source code and hardware available in the public domain or under the copyright that has been released under an open license

**Open Science infrastructures:** scientific equipment or sets of instruments, knowledge-based resources such as collections, repositories, archives and scientific data, open computational and digital infrastructures, needed to support Open Science and serve the needs of different communities

**Open engagement of societal actors:** citizen and participatory science and other extended collaboration between scientists and societal actors beyond the scientific community, opening up practices and tools that are part of the research cycle and by making the scientific process more inclusive and accessible to the broader inquiring society

**Open dialogue with other knowledge systems:** recognition of complementarities between diverse epistemologies, including indigenous knowledge systems





# COMMUNITY

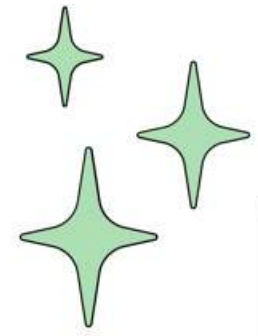
Peers, Colleagues, PhDs, Postdocs, Research Assistants, Lab Technicians, PIs,  
Supervisors, Research Heads, Masters Students, Undergrads.

Policy Makers, Industry people, Public

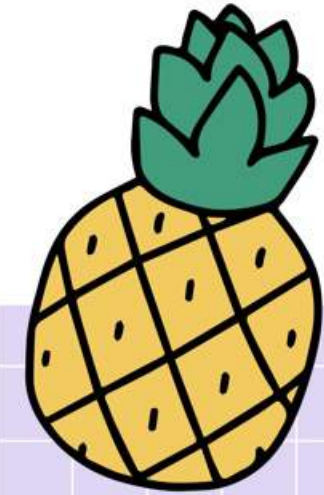
WHO IS THE COMMUNITY?







# PERSONAL PROFILE



## ORCID



uniquely identify authors and contributors of scholarly communication

## GOOGLE SCHOLAR



automatic aggregator of publications crawled from the web into a profile

## LINKEDIN



professional “social network” to connect, communicate, and disseminate

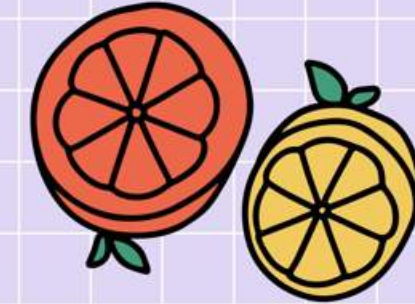
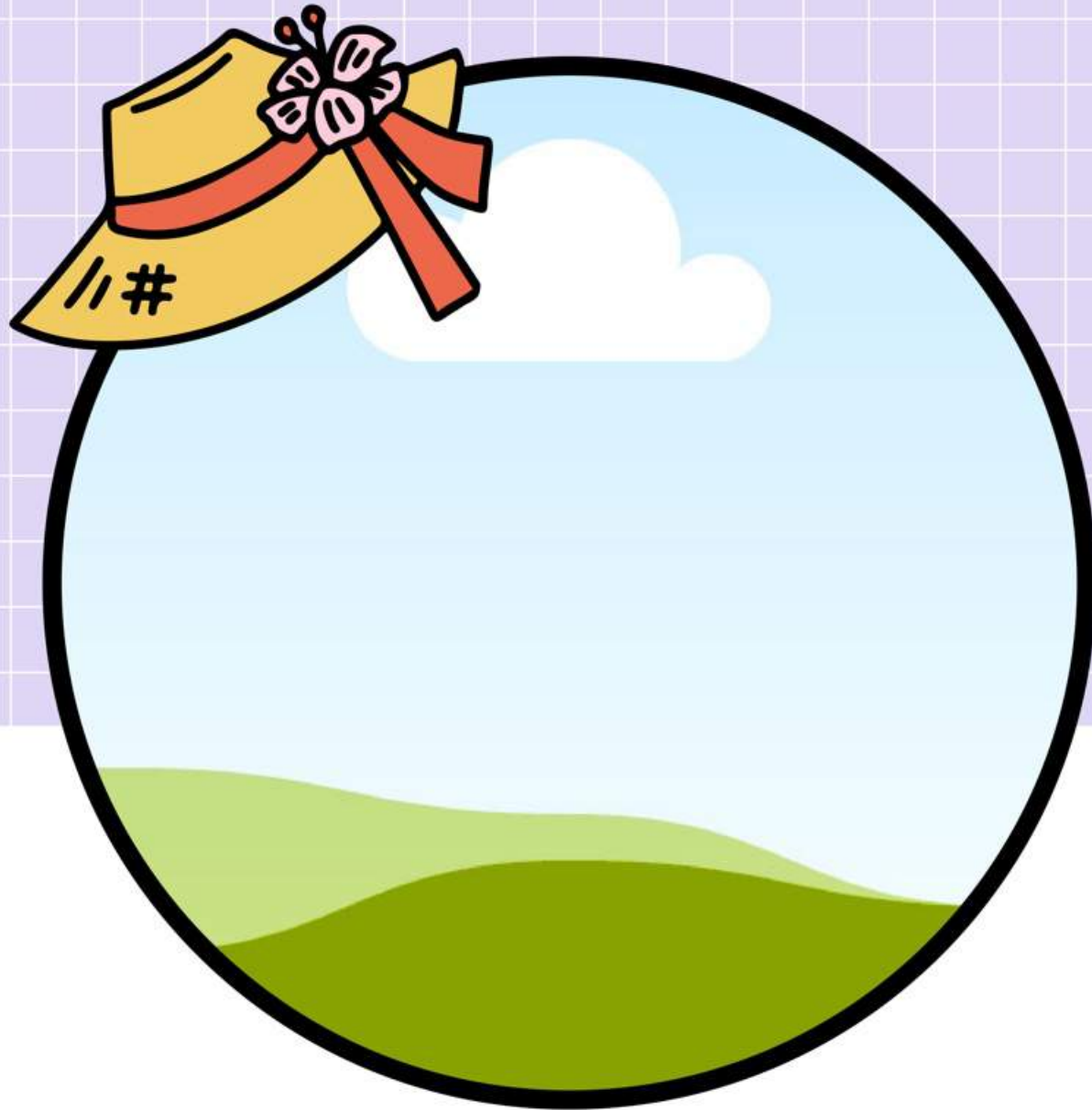
## PERSONAL WEBSITE



this is my page, there are many like it, but this one is mine



EXAMPLE.COM



# HELLO, IM JANUE DOE

I'm researching on this stuff, and other stuff, and then some more other stuff. I am interested in the areas of blah and blah. Here are my projects, and my publications, my data, and stuff.





# What information should be on this website?

The objective of this website is to:

1. Collect information related to our "regtech" work in a single location (i.e. this website)
2. Provide topical information collections e.g. work on GDPR, or work on AI Act
3. Share a comprehensive list of publications associated with "regtech"
4. Share outputs, methods, resources - anything that can be "reused" or "extended" to encourage adoption and collab
5. Share details about projects (funded as well as adhoc)
6. Share details about involvements in activities and groups e.g. DPVCG, NSAI, CEN, ISO
7. Share the people involved and show presence of a 'team'
8. Provide contact information



## RESEARCH

## # OVERVIEW

I am an Assistant Professor at the [School of Computing](#) and my research interests include application of semantics towards solving real-world problems in areas like data protection, legal and regulatory compliance, and consent management. I (along with my colleagues in Dublin) explored the application of linked data and semantic web technologies with a particular focus on consent and provenance. I chair the [Data Protection and Privacy Community Group \(DPVCG\)](#) – which develops interoperable standards for data protection activities based on legal and practical requirements. I am also a member of the [National Cyber Security Centre \(NCSC\)](#) Ireland (NSAI) and work on standardisation for [CEN/CENELEC](#).

You can see more information about my:

- [research interests](#)
- [publications](#)
- [draft publications](#) for feedback and collaboration
- [past and present research projects](#)
- [research activities, group memberships](#)
- [events and conference organisation](#)
- [participation in peer-review](#)
- [supervision and mentoring](#)
- [teaching, lectures, seminars](#)

[harshp.com/research](https://harshp.com/research)

## # RESEARCH INTERESTS

Privacy

I'm interested in the exploration of issues regarding privacy, especially those related to use of technology and data protection. My primary research interest is in exploring existing work in identification and authentication technologies and innovations.

Consent

Consent is a broad mechanism based on choice. My primary interest is in exploring protection issues, where it is used to manage research revolves around quantifying and identifying and mitigating issues.

General Data Protection Regulation

The GDPR and other recent data protection laws collecting, using, sharing personal data. My research interest revolves around exploring representation for evaluating compliance.

Semantic Web and Data Models

The Semantic Web standards, based on the web infrastructure and data relationships. My interest lies in exploring information from other research.

## # PUBLICATIONS

See full list of publications with links to copies and resources [here](#). Lists also available at [Google Scholar](#) and [dblp](#).

## # SELECTED PUBLICATIONS

To Be High-Risk, or Not To Be High-Risk: Applications and Harmonisation

Mon May 01 2023 *Confere*  
Conference on Fairness, Accountability, and Transparency  
*Delaram Golpayegani\*, Harshvardhan J. Pandit\**

Making Sense of Solid for Data Protection

Mon Feb 13 2023 *Journal*  
MDPI Information  
*Harshvardhan J. Pandit\**

Creating A Vocabulary for Data Protection

Fri Oct 11 2019 *Confere*  
International Conference on Semantic Web  
*Harshvardhan J. Pandit\*, Javier Fernandez, Ramisa Wenning*

## # PEER-REVIEW

## # Reviewer for Journals

- [Future Generation Computer Systems \(FGCS\)](#)
- [Journal of Data Protection & Privacy \(JDPP\)](#)
- [Journal of Information Security and Applications \(JISA\)](#)
- [Journal of Personal and Ubiquitous Computing \(JPUC\)](#)
- [Journal of Responsible Technology \(JRT\)](#)
- [Journal of Web Semantics \(JWS\)](#)
- [Semantic Web Journal \(SWJ\)](#)
- [IEEE Access \(IEEE Access\)](#)

## # Reviewer for Conferences/Workshop

- AICS (Irish Conference on Artificial Intelligence and Cognitive Science) [2022](#) [2018](#)
- CIKM (29th ACM International Conference on Information and Knowledge Management) [2022](#)
- CKG (Workshop on Contextualized Knowledge Graphs) [2019](#)
- DataValue (Workshop on Governing Value: The Practice of Exploiting Data Value) [2022](#)
- ESWC (Extended Semantic Web Conference) [2023](#) [2022](#) [2021](#) [2020](#) [2019](#) [2018](#)
- FAccT (Conference on Fairness, Accountability, and Transparency) [2023](#)
- ISWC (International Semantic Web Conference) [2023](#) [2022](#) [2021](#) [2020](#) [2019](#) [2018](#)
- IWPE (International Workshop on Privacy Engineering) [2022](#)
- JURIX (International Conference on Legal Knowledge and Information Systems) [2022](#)

# 05 – Community



# My workflow



## Create a new repository

A repository contains all project files, including the revision history. Already have a project repository elsewhere? [Import a repository](#).

*Required fields are marked with an asterisk (\*).*

Owner \*



Repository name \*

open-data-for-project

✓ open-data-for-project is available.

Great repository names are short and memorable. Need inspiration? How about **supreme-palm-tree** ?

Description (optional)

This repository provides access to data and code from the project.



**Public**

Anyone on the internet can see this repository. You choose who can commit.



**Private**

You choose who can see and commit to this repository.

## Initialize this repository with:



**Add a README file**

This is where you can write a long description

## Add .gitignore

.gitignore template: None ▾

Choose which files not to track from a list of temp

## Choose a license

License: Apache License 2.0 ▾

A license tells others what they can and can't do v

personal repo - Github  
ADAPT repo - Gitlab





# Open Science/Data = Permissive License

## Choose an open source license

An open source license protects contributors and users. Businesses and savvy developers won't touch a project without this protection.

{ Which of the following best describes your situation? }



**I need to work in a community.**

Use the [license preferred by the community](#) you're contributing to or depending on. Your project will fit right in.

If you have a dependency that doesn't have a license, ask its maintainers to [add a license](#).



**I want it simple and permissive.**

The [MIT License](#) is short and to the point. It lets people do almost anything they want with your project, like making and distributing closed source versions.

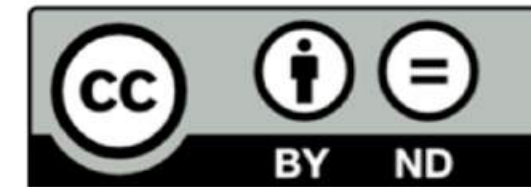
[Babel](#), [.NET](#), and [Rails](#) use the MIT License.



**I care about sharing improvements.**

The [GNU GPLv3](#) also lets people do almost anything they want with your project, *except* distributing closed source versions.

[Ansible](#), [Bash](#), and [GIMP](#) use the GNU GPLv3.





# My workflow

- [1. \*\*TODO\*\* Submitting the Paper: \[0/8\]](#)
- [2. \*\*TODO\*\* Reviews received: \[0/5\]](#)
- [3. \*\*TODO\*\* Camera-ready: \[0/2\]](#)
  - [3.1. \*\*TODO\*\* Submit Camera-Ready version \[0/4\]](#)
  - [3.2. \*\*TODO\*\* Upload pre-print](#)
    - [3.2.1. \*\*TODO\*\* Upload pre-print to Zenodo](#)
    - [3.2.2. \*\*TODO\*\* Upload pre-print to TCD TARA](#)
- [4. \*\*TODO\*\* Presentation: \[0/2\]](#)
- [5. \*\*TODO\*\* Published: \[0/1\]](#)

## 1. **TODO** Submitting the Paper: [0/8]

- [ ] Check format and page requirements
- [ ] Check single/double blind requirements
- [ ] Acknowledgements for funding
- [ ] Email addresses
- [ ] OA link in abstract
- [ ] Link to resources
- [ ] Submit paper
- [ ] Save copy as submitted version

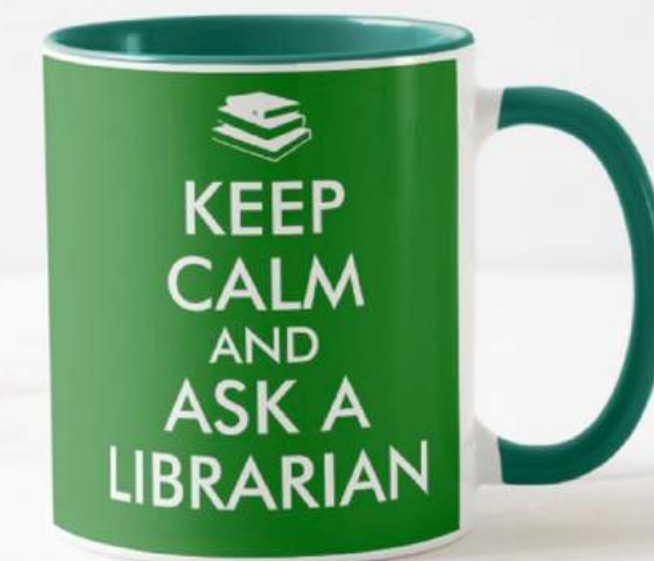
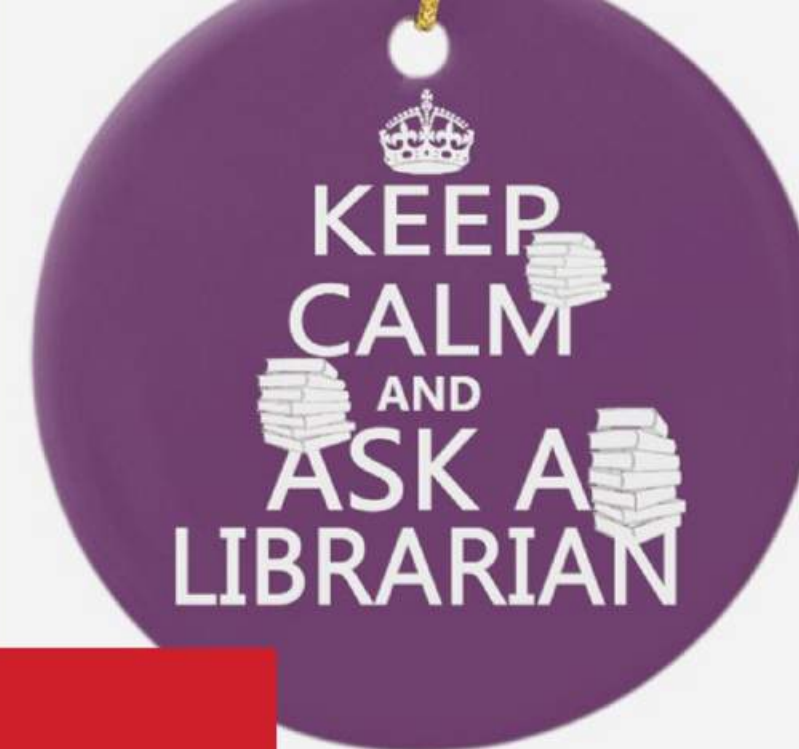
## 5. **TODO** Published: [0/1]

- [ ] Update trackers
  - [ ] Theme-E
  - [ ] harshp.com website
  - [ ] TCD/RSS





**KEEP  
CALM  
AND  
ASK A  
LIBRARIAN**





# Thank You

Harshvardhan Pandit

[harshvardhan.pandit@adaptcentre.ie](mailto:harshvardhan.pandit@adaptcentre.ie)

<https://harshp.com/research/presentations>

