

# A staple for researchers. Here's the catch.



# Git is a culture, not a tool

- 42% use Git-based platform as **storage** (!)
- 51% version their **code** with Git\* platforms
- 27% use Git also for versioning of **data**



[10.5281/zenodo.18416932](https://doi.org/10.5281/zenodo.18416932)

# So, why those asterisks?

## ERC Open Data and Code Guidelines

| DATA TYPE                       | WHERE TO SUBMIT*                                                                                                | WHAT TO INCLUDE IN THE DATA AVAILABILITY SECTION OF YOUR ARTICLE |
|---------------------------------|-----------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------|
| Latest source code              | <a href="#">GitHub</a> , <a href="#">BitBucket</a> , <a href="#">SourceForge</a> or <a href="#">Google Code</a> | URL                                                              |
| Archived source code            | <a href="#">Zenodo</a>                                                                                          | Title, DOI and license* used                                     |
| Deposits of mixed data and code | <a href="#">Code Ocean</a>                                                                                      | Title, DOI, embed code for interactive reanalysis tool           |
| Software                        | Authors may host software where they wish, though it is strongly recommended to use a stable URL                | URL                                                              |

\*\* Github is a commercial platform that is well-suited to code development, but it is not a data archiving tool. Therefore, a copy of the code has to be archived in a data repository. For instance, you can set up a [connection between Zenodo and Github](#)  to ensure a permanent record of your code and make it citable.

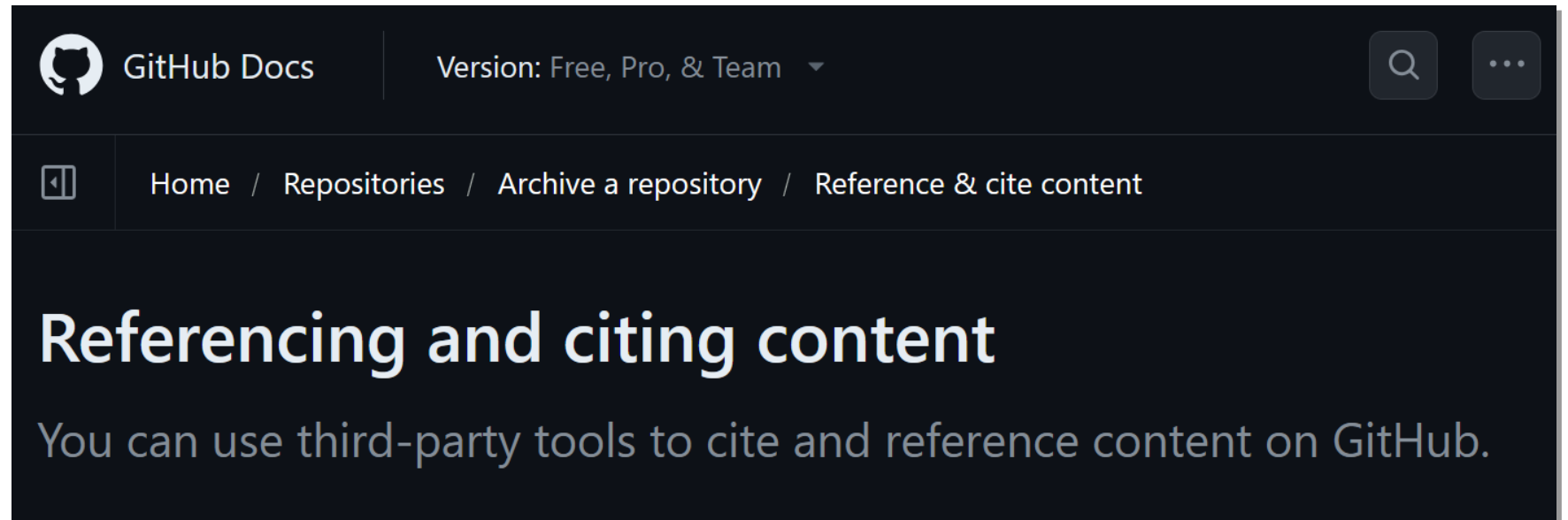
SNSF Open Research Data repositories

# Sharing ≠ Publishing

- License clarity not enforced
- Repository content is mutable
- No standardized discovery metadata
- No long-term preservation commitment
- No persistent identifier (DOI or other PID)
- Account automatically tied to existing MS account
- Repos / Accounts removal under political pressure
- AI training regardless of license (since 2021, allegedly)

**Not adapted for "publishing" nor  
for “archiving” academic output**

# GitHub says ± the same



## Issuing a persistent identifier for your repository with Zenodo

To make your repositories easier to reference in academic literature, you can create persistent identifiers, also known as Digital Object Identifiers (DOIs). You can use the data archiving tool [Zenodo](https://zenodo.org/) to archive a repository on GitHub and issue a DOI for the archive.

# But GitHub also says ...

[copilot.github.com/trust/page/faq](https://copilot.github.com/trust/page/faq)

Does GitHub use Copilot Business or Enterprise data to train AI models?

Copy link

GitHub does not use Copilot Business or Enterprise customer data to train AI models. Business and Enterprise customers' data is covered by GitHub's Data Protection Agreement, which prohibits such use without customer authorization.

[github.blog](https://github.blog)

Blog

Home / News & insights / Company news

## Updates to GitHub Copilot interaction data usage policy

From April 24 onward, interaction data—specifically inputs, outputs, code snippets, and associated context—from Copilot Free, Pro, and Pro+ users will be used to train and improve our AI models unless they opt out.

## GitHub and Trade Controls

GitHub.com, GitHub Enterprise Server, and the information you upload to either product may be subject to trade control regulations, including under the U.S. Export Administration Regulations (the EAR).

[GitHub and Trade Controls](#)

[github-terms](#)

### 4. License Grant to Us

You grant GitHub and our Affiliates the right to store, host, archive, parse, display, and make copies of Your Content as necessary to provide, develop, and improve the Service, including by training AI Features, and for the purpose of training, developing, and improving artificial intelligence and machine learning models and technologies of our Affiliates. This license includes the right to do things like copy it to our database and make backups; show it to you and other Users; parse it into

# ... and impacts entire countries

**ZD  
NET**

**GitHub starts blocking developers in countries facing US trade sanctions**

“If you use GitHub's online services in a country facing US sanctions, you could be about to be kicked off all but the most basic offerings” ([ZDNET](#))

**InfoWorld**

**Judge dismisses lawsuit over GitHub Copilot AI coding assistant**

“GitHub, its owner Microsoft and OpenAI free to train on code samples despite DMCA” ([Infoworld](#))

**TC TechCrunch**

**GitHub confirms it has blocked developers in Iran, Syria and Crimea**

“The impact of U.S. trade restrictions is trickling down to the developer community” ([Techcrunch](#))

**IEEE Xplore®**

**The Impact of Sanctions on GitHub Developers and Activities**

“complying with international policies while maintaining a global community” ([IEEE article](#))

**The Register**

**GitHub courts controversy by suspending Tornado Cash developers and reneging on cookie commitments**

“If you're looking for free speech or privacy, move along” ([The Register](#))



# OSS discovery: a quest

- [CHAOSS](#) metrics (Linux Foundation®)
- [OSS Insight](#) explorer (TiDB Cloud)
- [French Open Science Monitor](#) (France)
- [PlumX Metrics](#) (Elsevier)
- [OSPO-RADAR](#) (Software Heritage)
- [International Standard for Software Artifact Identification](#) (ISO/IEC 18670:2025)
- [“How to Value Open Source Contributions?”](#) (CERN)
- [Towards Open Bibliometric Indicators – TOBI](#) (ETHZ Library)
- [Open Pulse](#) (EPFL / SDSC)
- [Openness Score](#) (FHGR)
- ...
- openness.switch.ch → opendatanavigator.switch.ch → resources.eosc.ch → ???

# Openness is not trendy anymore



HBS AI Content &amp; Learning



HBS AI Affiliate Research

## Revealing Value: The Economic Power of Open Source Software

AI &amp; Future of Business Technology | Data, Privacy, Security, and Regulation | Future of Work | Human-Centered Technology |

New Operating Models and Business Models | Performance and Metrics

“As they consider the stakes of the research, the authors draw a parallel with the concept of the *tragedy of the commons*” ([Harvard Business School](#))

## swissuniversities

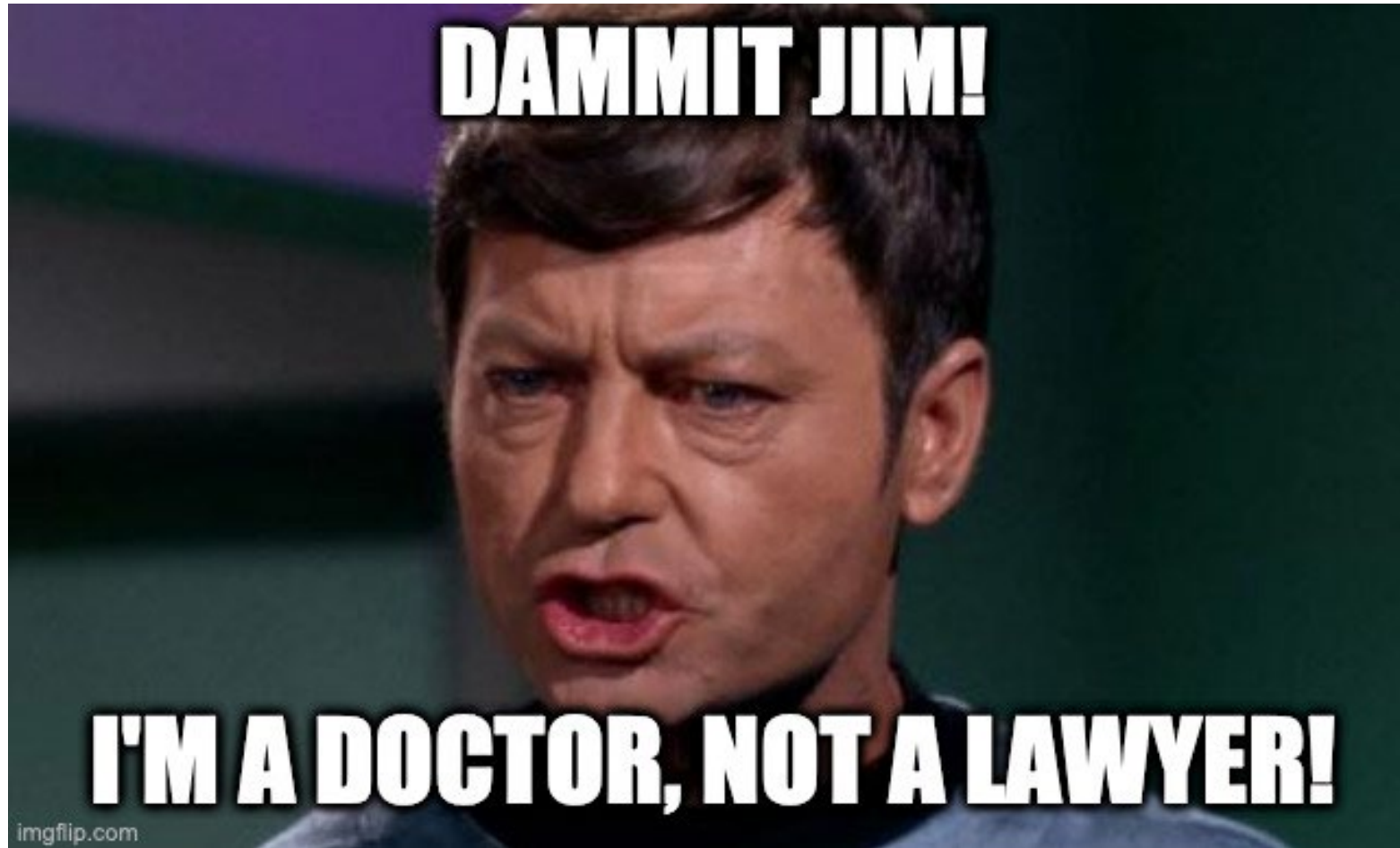
swissuniversities  
Effingerstrasse 15, Case Postale  
3001 Berne  
[www.swissuniversities.ch](http://www.swissuniversities.ch)

## Knowledge Security in Switzerland: A Strategic Framework for Higher Education Institutions and Authorities

“While research integrity is primarily concerned with ethical conduct and scientific rigor, knowledge security is about protecting valuable research assets and maintaining competitive advantages or national security interests. These two concepts can sometimes create tensions” ([Swissuniversities](#))

# Are researchers... criminals?

- ⊗ No institutional, cantonal, or federal rule explicitly requires using an institutional email when creating GitHub/GitLab accounts for project work
- ⊗ The **employer alone** can exercise the **exclusive rights of use** of a computer program created under an employment contract (Copyright Act, art. 17)
- ⊗ **Authorship** and moral rights (ex., to be named) stay with individual researchers; **economic/exploitation rights** for IP generated while working at an institution belong to the institution
- ⊗ **Failing to disclose** the use of GitHub is a breach of internal institutional expectations or employment-related obligations, but it's not punishable, so ... they are **not criminals**



| Measures                                                             | Mechanisms                                                            | Effort by Researcher | Effort by Institution |
|----------------------------------------------------------------------|-----------------------------------------------------------------------|----------------------|-----------------------|
| Harvest code/data availability statements from publications          | Textmine metadata for repo URLs (one-time build)                      | None                 | High                  |
| CITATION.cff / codemeta.json / README.md / howfairis config          | Template files and actions incl. ORCID/ROR, that facilitate discovery | Low                  | Low                   |
| GitHub → Zenodo + Software Heritage                                  | Promote and mint Zenodo, inherit ORCID/ROR                            | Low                  | High                  |
| Update DMP with info on code repos (ex. platform, repo URL, license) | Add mandatory fields and final check                                  | Low                  | Medium                |
| Support GitLab Community Ed. as replacement                          | Provide institutional platform, and add minting pipeline              | Low                  | High                  |
| Annual software inventory prompt                                     | Short survey to PIs on repos (one-time build)                         | Low                  | Medium                |
| PI-level disclosure accountability                                   | Add “code repos” to projects’ reporting                               | Low                  | Low                   |
| SWHID mining for institutional attribution                           | Trace code using Software Heritage graph                              | None                 | High                  |

# 1st measure

**Mandate PIs** to use & to make their  
researchers use their institutional  
**@** email address to signup  
on  external services

# Take-home messages

- ✓ Open code on GitHub is a fantastic **starting point**, not an end, because sharing ≠ publishing
- ✓ Code that is FAIR would be more easily treated **as academic output** (legal, bibliometrics, etc.)
- ✓ Rather than chasing OSS in the wild, institutions can adopt measures to **guide researchers**
- ✓ Instead of fragmented pet projects, HEIs can leverage the existing **Git culture** of researchers

# Take-home message = Let's not reinvent the wheel

- ☑ Open code on GitHub is a fantastic **starting point**, not an end, because sharing ≠ publishing
- ☑ Code that is FAIR would be more easily treated **as academic output** (legal, bibliometrics, etc.)
- ☑ Rather than chasing OSS in the wild, institutions can adopt measures to **guide researchers**
- ☑ Instead of fragmented pet projects, HEIs can leverage the existing **Git culture** of researchers

# EPFL

OPEN SCIENCE DAYS 2026

# Thank you for your attention!

■ École  
polytechnique  
fédérale  
de Lausanne

CC-BY 4.0  

PRESENTED BY

**Francesco Varrato**  
Research Data Management  
Specialist - EPFL Library