

**EOLE 2024**

**15<sup>th</sup> edition**

# Navigating the New EU Regulatory Landscape in Open Source

**29 November 2024**

**Turin**



# 4 – Open Source and Open Science

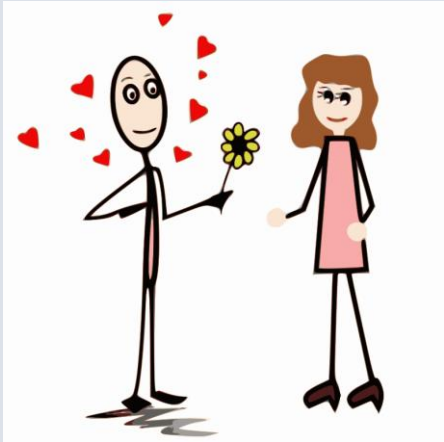
---

16:00 – 17:30

## ***Speakers***

- ▣ *Ludovica Paseri, University of Torino*
- ▣ *Malcolm Bain, Across Legal*
- ▣ *Aurelio Ruiz, UPF*

# Open Source and Open Science



## A happy partnership?

Malcolm Bain

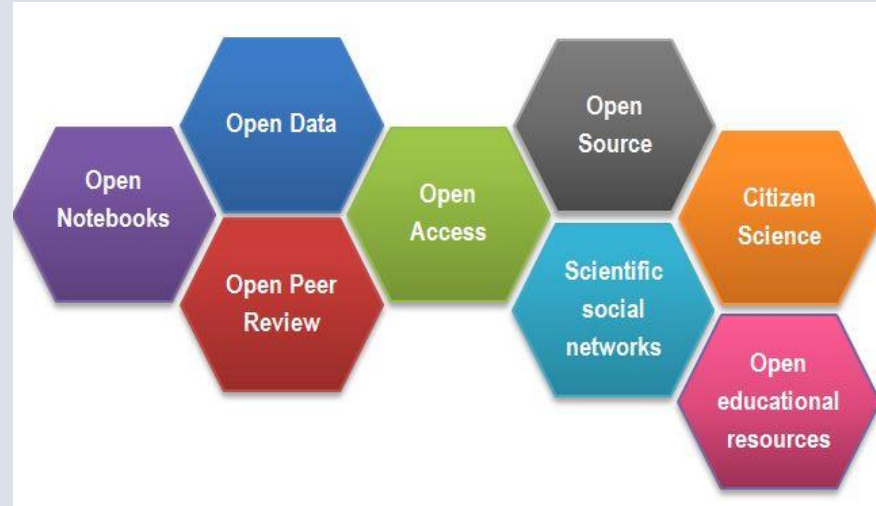
Across Legal, FSFE

# Can Science learn from Open Source?

# What can Science learn from Open Source?

# Open Science

**Umbrella term** that involves various movements aiming to remove the barriers for **sharing any kind of output, resources, methods or tools, at any stage of the research process**

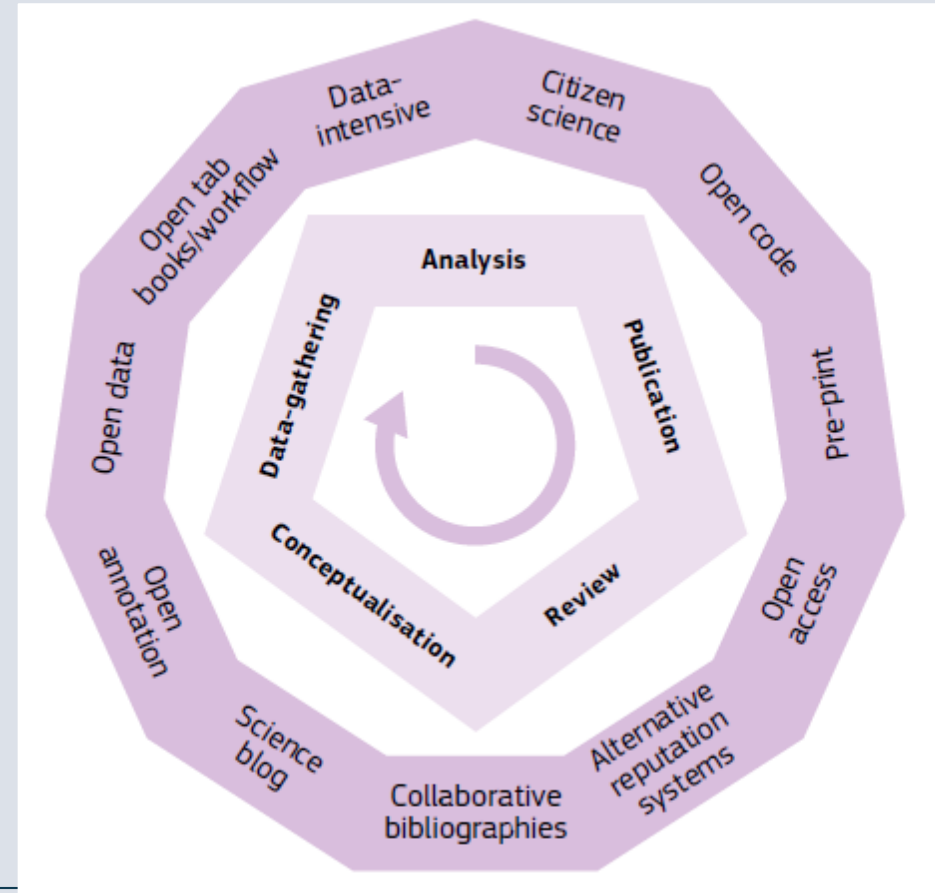


.Objective: to make the primary outputs of **research results** – publications and the research data – **publicly accessible** in **digital format** with **no or minimal restriction**

.Principal focus is on **Open Research Data** and **Open Access to (scientific) publications**

# Open Science as a process

Open Science is seen as a “process”: being open in all the stages of research, from hypothesis, data collection, analysis, to publication, preservation and reuse ...



# Objectives of Open Science

Open science aims to ensure the **free availability and usability** of scholarly **publications**, the **data** that result from scholarly research, and the **methodologies**, including **code or algorithms**, that were used to generate those data.

*Open Science by Design- Realizing a Vision for 21st Century Research (2018),  
National Academies of Sciences, Engineering, and Medicine;*



# Why Do Open Science?



## Accessibility

- .Citizen science initiatives and engagement
- .Lifesaving access to medical and scientific information
- .Democratization of the scientific process
- .Increased Earth observation accessibility

## Efficiency & Innovation

- .Cleaner, more secure code with more contributors
- .More long-term maintenance assistance from the community
- .New monetized offices and data centers
- .Financially sound, streamlined, and transparent research spending

## Reproducibility

- .Increased transparency of results from research
- .More reliable results with confirmation of previous findings
- .Less pressure to come up with 'exciting' research in order to be published
- .More robust products of the scientific method

## Diversity & Inclusion

- .International accessibility
- .Breakdown of some systemic financial and funding barriers
- .Diversity among researchers and scientists
- .Equitable distribution of opportunity

# Doesn't this look a lot like Open Source?

“The overarching principle of open science by design is that research conducted openly and transparently leads to better science.” (National Academies of Sciences, Engineering, and Medicine. 2018. Open Science by Design: Realizing a Vision for 21st Century Research)



•The overarching principle of **open source** is that **software development** conducted openly and transparently leads to better **software**. (paraphrasing Eric Raymond, Cathedral and Bazaar, 1999)

# Open Science - Open Source: Similarities

## Nature: Software and Scientific Knowledge are

- Non rival
- Non scarce

## Objectives: Open software and science aim for

- Accessibility
- Reproducibility
- Reusability
- Replicability

Open Source = open knowledge = result of **an** open scientific process

**BUT...**

# Differences: Practices

## Open source

- Publish often
- Publish early
- Post publication peer review
- Correct errors, collaboratively
- Repeat

## Open Science

- Publish once, carefully
- Prior publication peer-review
- “Counter” papers to refute ideas / results, competitively?
- Repeat? Not necessarily

# Differences: Search for Protection

## Science

IPR objective: Strong protection - Patent (or similar), Designs, ...automatic copyright

-Examples: RSA, Google Page Rank, MP3 and MP4 compression... 5G protocols

IPR logic for investing in Science

-Monopoly for exploitation and further investment

## Open Source

Automatic protection by copyright

Open Source has to overturn the IPR (copyleft) to make it work for free and open source.

IPR logic for open source

- Ensure attribution and continued openness of the code

.

# Differences: Motivation

## Scientific motivations – institutional and personal

- Scientific reputation – publish!
- Keep it confidential until published (first to publish)
- Patent - Monopoly protection for licensing opportunities
- ROI in terms of monetary return and re-investment in R+D
- Impact measured in citations and ...  
\$\$\$\$ in licensing terms

## Open source

- Technical reputation, quality of code
- Publish early, share and improve ... (and maintain?)
- Impact measured in downloads, use and reuse, contribution to other projects

# Differences: **Implementation**

## Science

Implementation of many sciences requires significant infrastructure

## Open Source

Software = implementation / near zero marginal cost

.Exceptions – e.g. AI requires significant infrastructure



# Differences: Risk factors

.Covid 19 comment:

-Openly sharing results that have not yet been peer-reviewed can be very damaging if the media and the public take these findings at face value.


-while peer-review has a key role in maintaining quality, transparency and reproducibility of published articles, it is not sufficient to avoid the publication of flawed studies, and their use by the media and health authorities

-some of the misuses and abuses that we have highlighted are a direct result of the current metric-centered evaluation of research and researchers which has already been shown to lead to questionable research practices in the past

# Uncertainty for open science

Significant questions for science projects that want to be open:

- which assets (results) can be published and shared?
- which processes may be useful or required for opening and sharing with a community
- which community – is it the whole world or a more restricted set of privileged partners?



# Are Open Source and Open Science good bedfellows, and can Science learn from Open Source?

Some partial conclusions

# Growing importance of Software in Science

There is also a growing realization that as **scientific research depends more and more heavily on computer code** for simulations, calculations, analysis, visualization, and general data processing it is **important to have access to this code** just as it has traditionally been important to show (and derive) any new mathematical techniques introduced for analysis.

(Marcus Hanwell, [opensource.com](https://opensource.com))

# How to make open science work?

## What (open) science can learn from open source :

- Strength through openness and freedom
- Licensing establishes the ground rules of the communities
- Community and Governance are critical

## Need to

- Change approach to scientific incentives, expectations and reward
- Change assessment of impact of research – stop measuring patents and licensing revenue!
- One size does not fit all

## It depends

- Differences between basic and applied research
- Differences between digital / non digital science

# Practical implications

## More and more open source is being produced in R+D

- Publications should publish open code and open data “better” (e.g. with a license!)

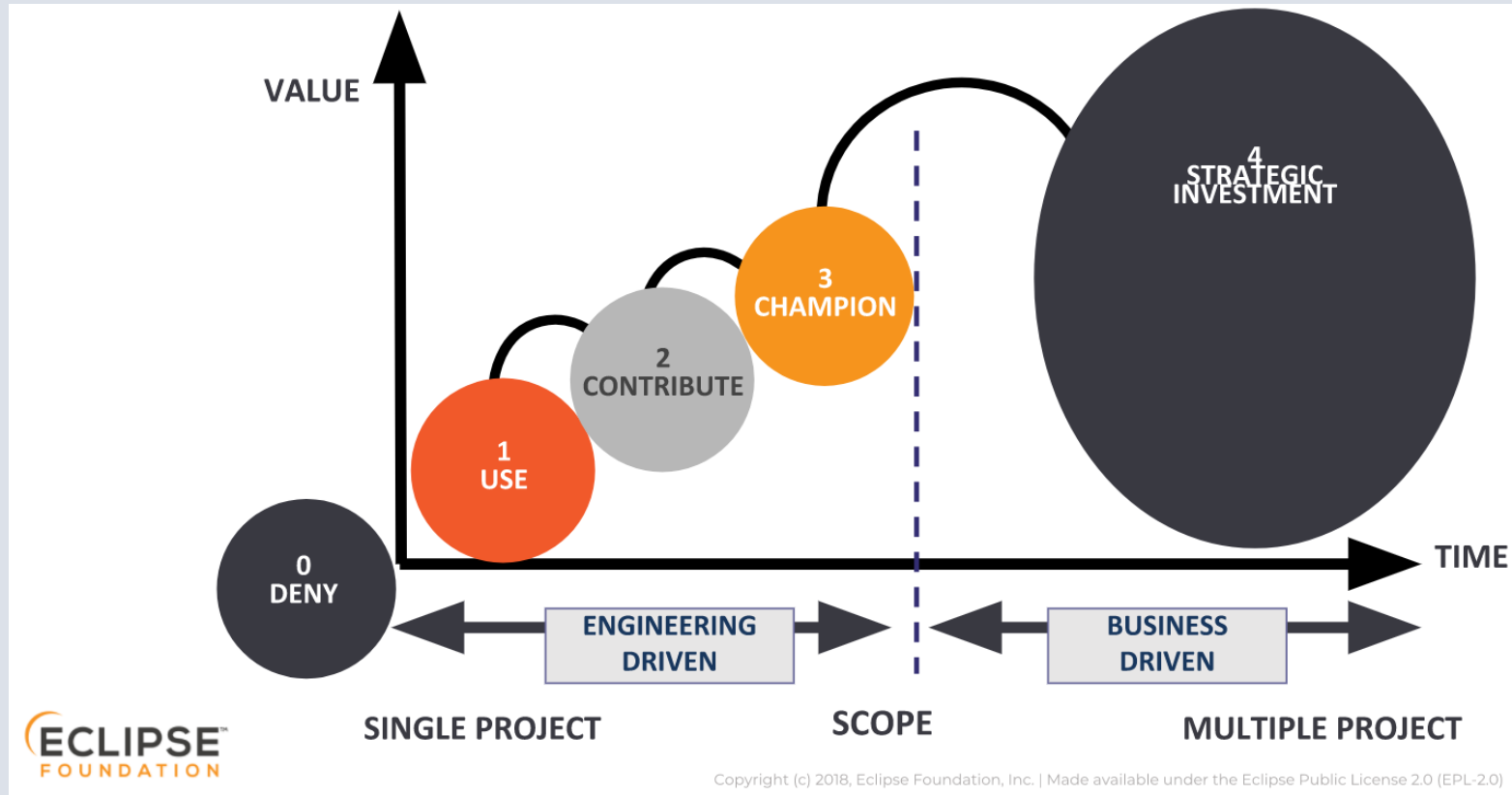
## Open source best practices should be (or could be better) implemented in research

- Open collaboration (environments, practices, incentives)
- License compliance procedures
- Governance processes
- Community management and sustainability

## Avoid pitfalls of non-openness:

- “academic non-commercial” licenses? (context specific)
- Non (open) data – particularly in ML based research
- Register to get access... / Access Agreements

# Open source adoption/maturity curve



Copyright (c) 2018, Eclipse Foundation, Inc. | Made available under the Eclipse Public License 2.0 (EPL-2.0)

# “Recommendation One”

“Research institutions should work to create a culture that actively supports **Open Science by Design** by better rewarding and supporting researchers engaged in open science practices.

Research funders should provide explicit and consistent support for practices and approaches that facilitate this shift in culture and incentives.”

National Academies of Sciences, Engineering, and Medicine. 2018. Open Science by Design: Realizing a Vision for 21st Century Research. Washington, DC: The National Academies Press.  
<https://doi.org/10.17226/25116>.



At its core open science is about getting back to those core values instilled by some of the earliest scientists that we should take no one's word for it, that it is essential all elements pertinent to a claimed discovery are published so that the results can be repeated and validated.

“Nullius in verba”

**Marcus  
Hanwell**



# Linus Torvalds

I often compare open source to science. To where science took this whole notion of developing ideas in the open and improving on other peoples' ideas and making it into what science is today and the incredible advances that we have had. And I compare that to witchcraft and alchemy, where openness was something you didn't do.

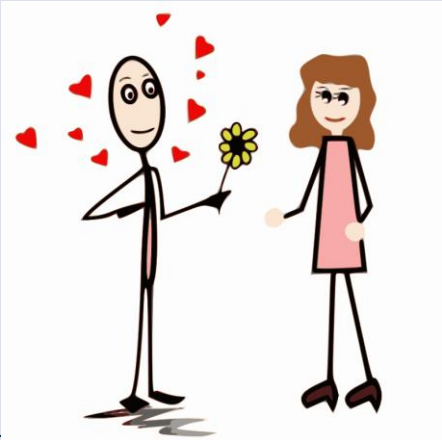


So... question for continuing the debate

Can Science (re)learn from Open Source...

... what open source learnt originally from science?

# Open Source and Open Science A happy partnership?



Malcolm Bain  
Across Legal, FSFE

[mbain@acrosslegal.com](mailto:mbain@acrosslegal.com)