

Beyond the Repository

Best practices for open source ecosystems
researchers

julia ferraioli and amanda casari
FOSDEM 2024



bit.ly/beyond-the-repo-fosdem24

julia ferraioli

AI/ML Open Source Strategist at AWS

Education: Master of Science in Computer Science

Research interests: motivations, standards, and collaboration patterns in open source software

juliaferraioli.com | opensourcestories.org



amanda casari

Researcher & Engineer, Google Open Source Programs Office

Education: Master of Science in Electrical Engineering

Research interests: open source as complex ecosystems and networks, data science + ethics with a feminist lens, human-in-the-loop systems design

@amcasari | opensourcestories.org



Juniper Lovato (in absentia)

Research Assistant Professor, Computer Science at UVM

Education: PhD in Complex Systems & Data Science

Research interests: tech ethics, complex systems,

computational social science, science of stories

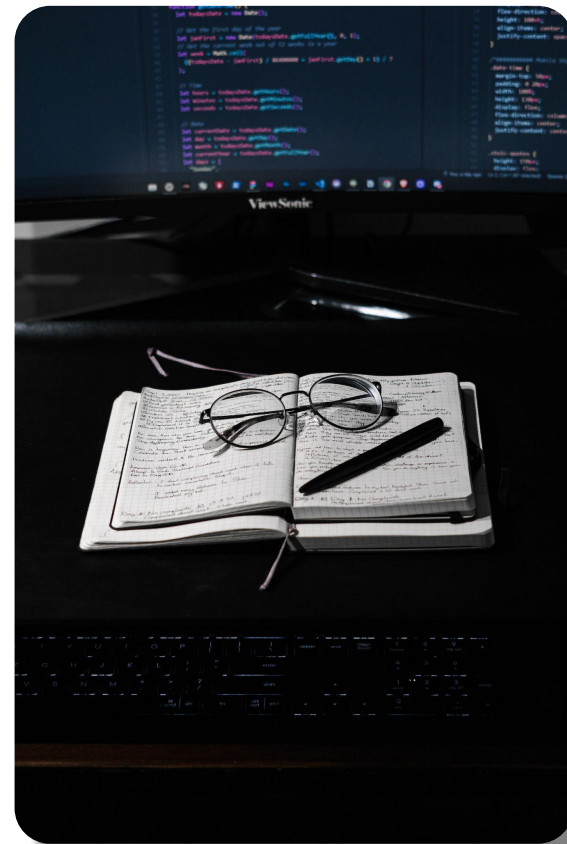
juniperlovato.com | [@juniperlov](https://twitter.com/juniperlov)





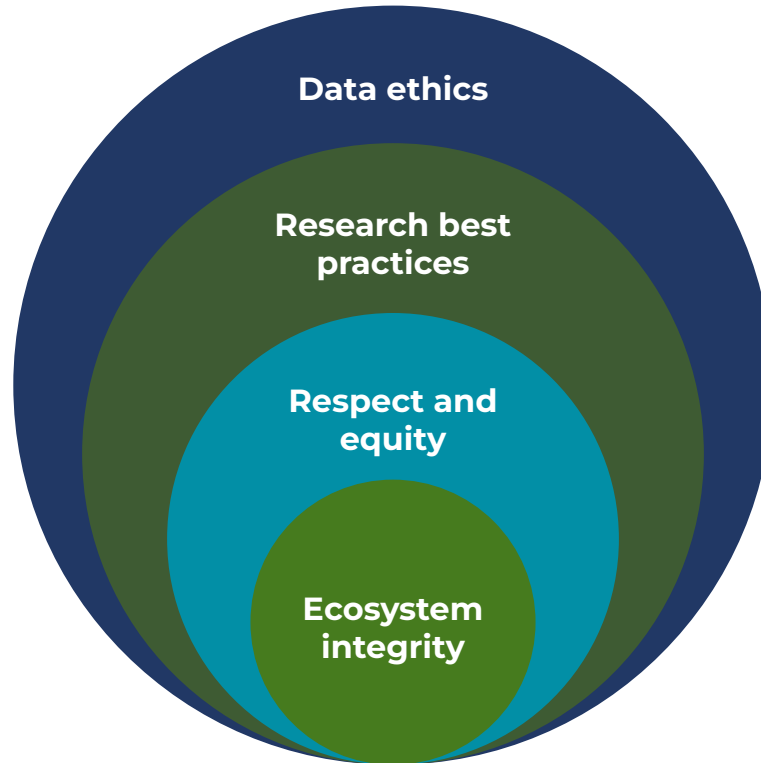
Research

Opportunity to
bridge the gap



Practice

**When researching open source ecosystems,
consider...**

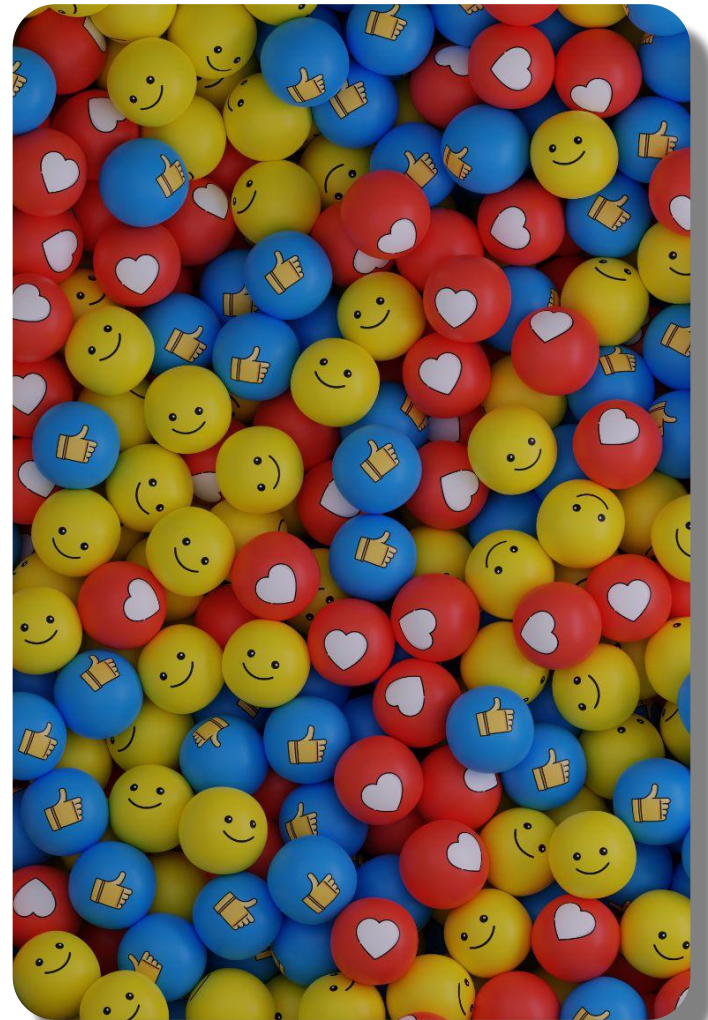


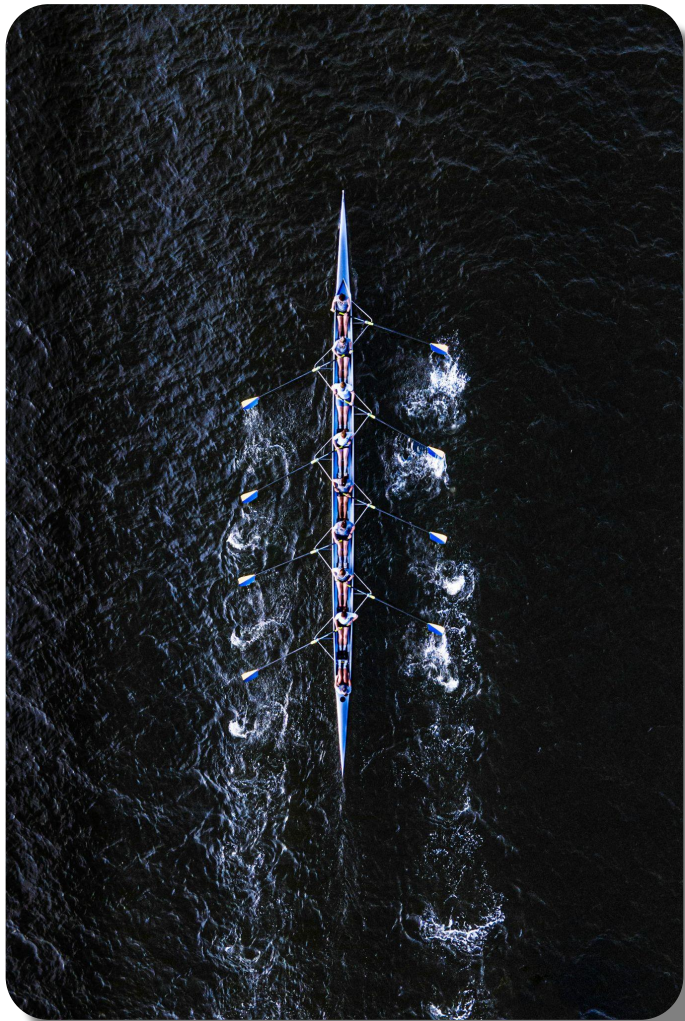
Data ethics



"Summary: We found no general model to predict software developer's well-being from software repositories."

– bit.ly/repo-negative-result



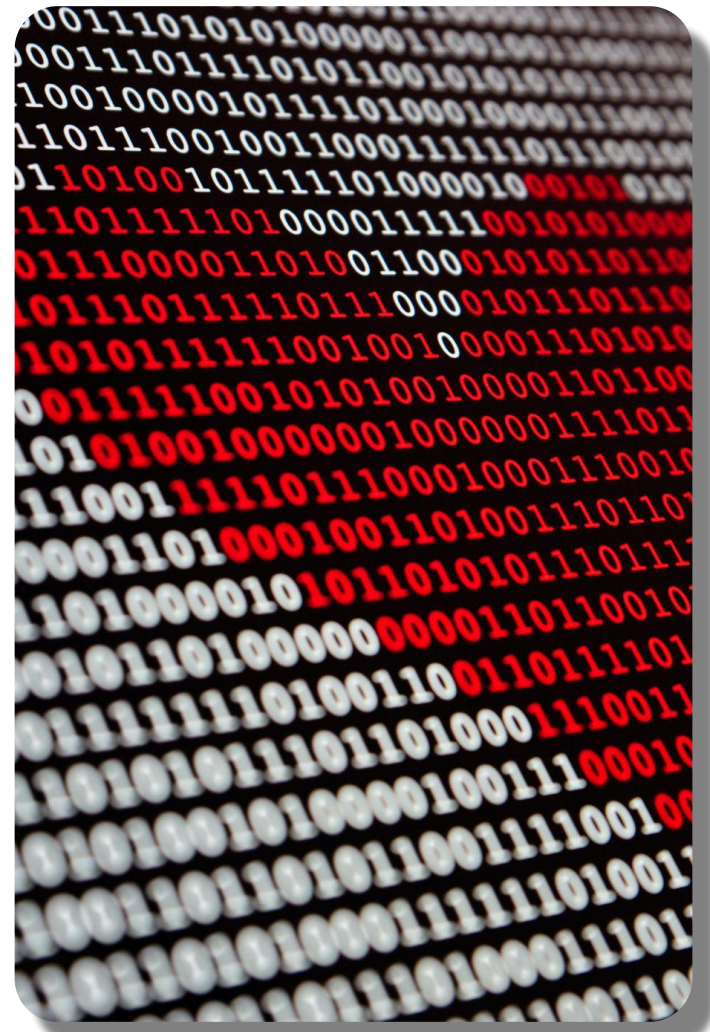


**Retain the "socio"
element when scoping
research questions of
a sociotechnical
ecosystem.**

Research best practices

"Contextual integrity ties adequate protection for privacy to norms of specific contexts, demanding that information gathering and dissemination be appropriate to that context and obey the governing norms of distribution within it."

– *Helen Nissenbaum*
(bit.ly/privacy-integrity)



"...there's a difference between people browsing my profile and deciding to ping me and people sending emails to tens of thousands of emails because they have one convenient data set..."

– *GitHub User, 2016*
(bit.ly/ghtorrent-32)





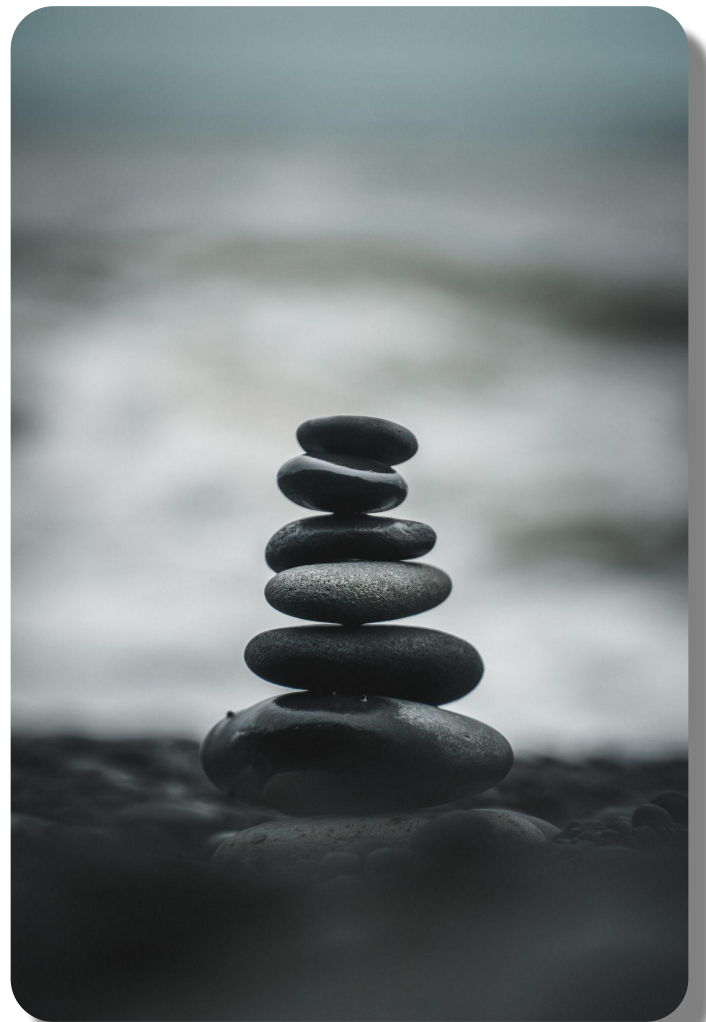
Openness isn't a good in itself, find the balance between privacy, ethics, and transparency when processing and sharing data.

Respect and equity



"I've already told my partner that, if and when we decide to start having kids, I will probably quit open source for good. I just can't see how I'll be able to make the time for both raising a family and doing open source."

– *Lawson, 2017*
(bit.ly/family-and-oss)





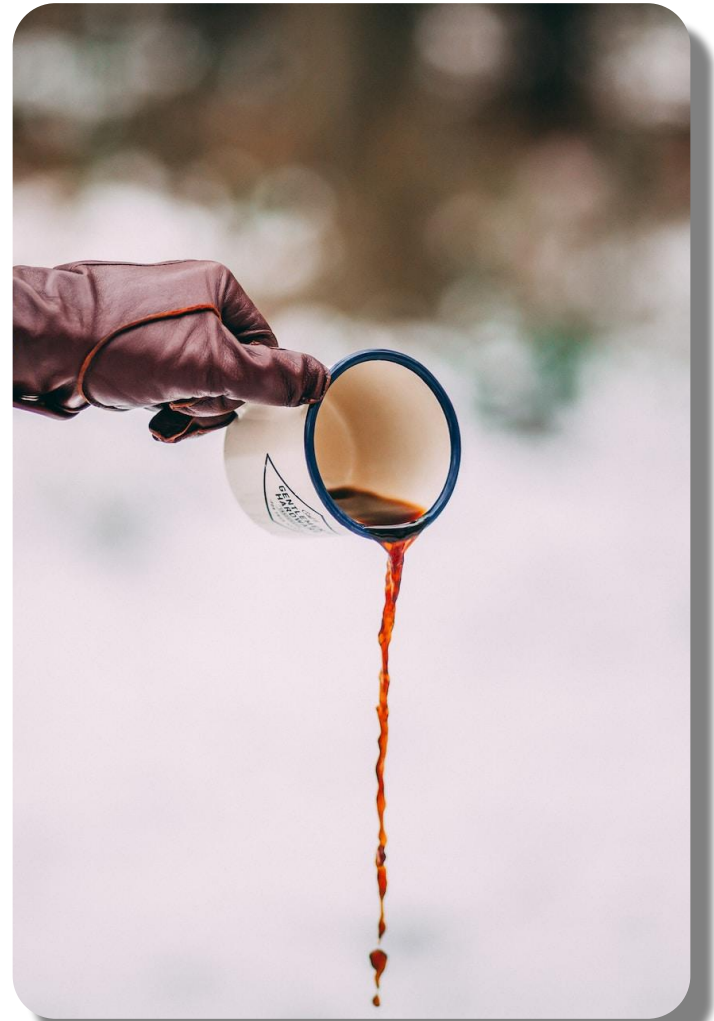
Assume that the economic incentives and availability of the people who keep the lights on are not evenly distributed.

Ecosystem integrity



"The researchers crossed a line they shouldn't have crossed. Nobody hired this group. They just chose to do it. And a whole lot of people spent a whole lot of time evaluating their patches."

– *Greg Scott, 2021*
(bit.ly/banned-from-linux)





Always treat open source ecosystems as systems "in production."

Actively work with open source ecosystems.

Consider the ramifications of research.

Look beyond the repository.



We can bridge the gap between researchers
and practitioners by working together.



Thank you! — References

- **These slides:** bit.ly/beyond-the-repo-fosdem24
- Beyond the Repository: bit.ly/beyond-the-repository
- The CARE Principles for Indigenous Data Governance: bit.ly/CARE-principles
- The FAIR Guiding Principles for scientific data management and stewardship: bit.ly/FAIR-principles
- Why Do People Give up FLOSSING? A Study of Contributor Disengagement in Open Source: bit.ly/giving-up-flossing
- On the Feasibility of Stealthily Introducing Vulnerabilities in Open-Source Software via Hypocrite Commits: bit.ly/hypocrite-commits
- Individual differences limit predicting well-being and productivity using software repositories: a longitudinal industrial study: bit.ly/repo-negative-result