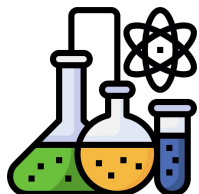


# Understanding the Sustainability Challenges for Building Open-Source Scientific Software

**Shurui Zhou**

Assistant Professor

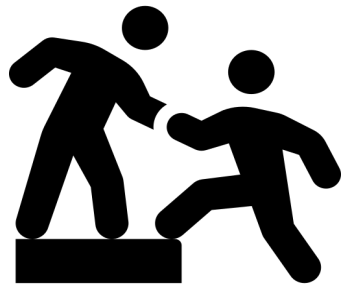
 @shuishuiblue



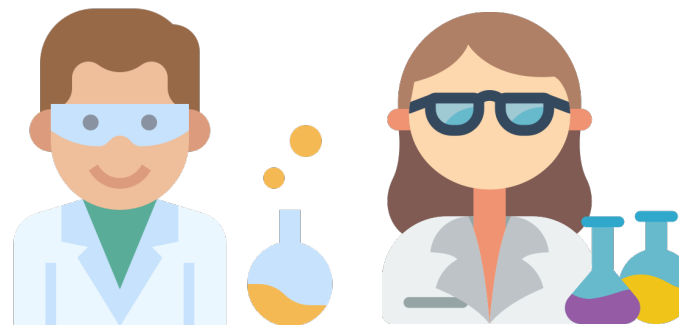
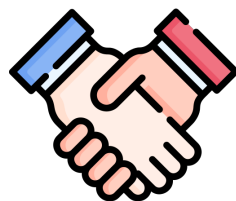
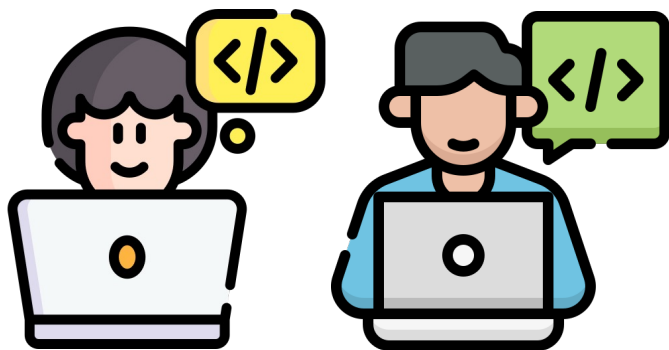
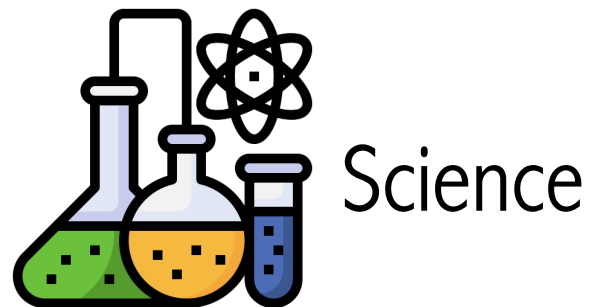
UNIVERSITY OF  
**TORONTO**

# Sustainability Challenges in Open-Source

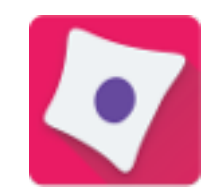
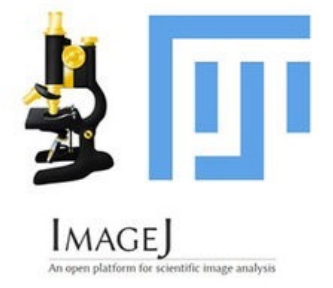
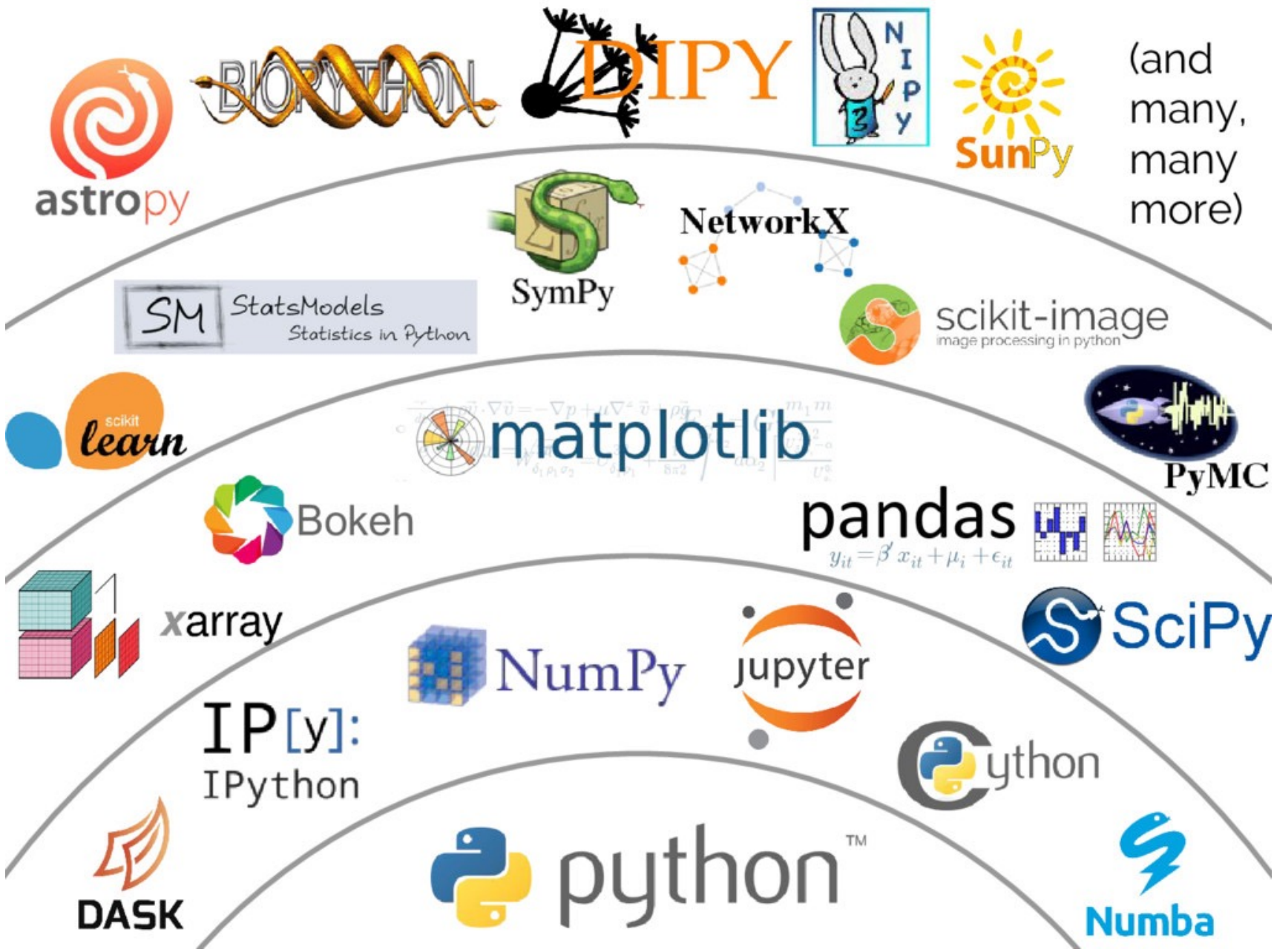
- Sustaining the project
- Sustaining the community



# Open-Source Scientific Software

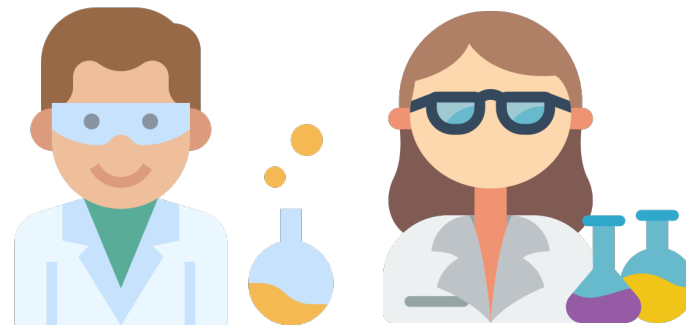
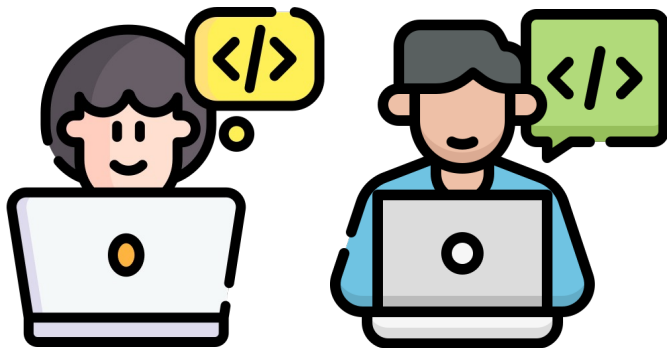
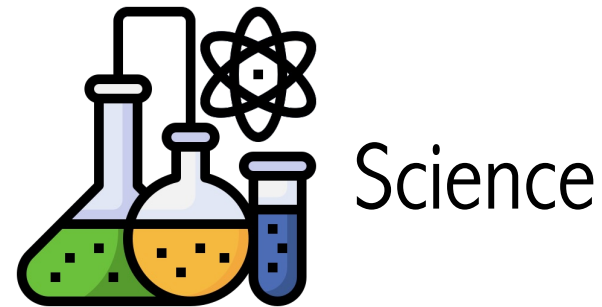


**Interdisciplinary collaboration**



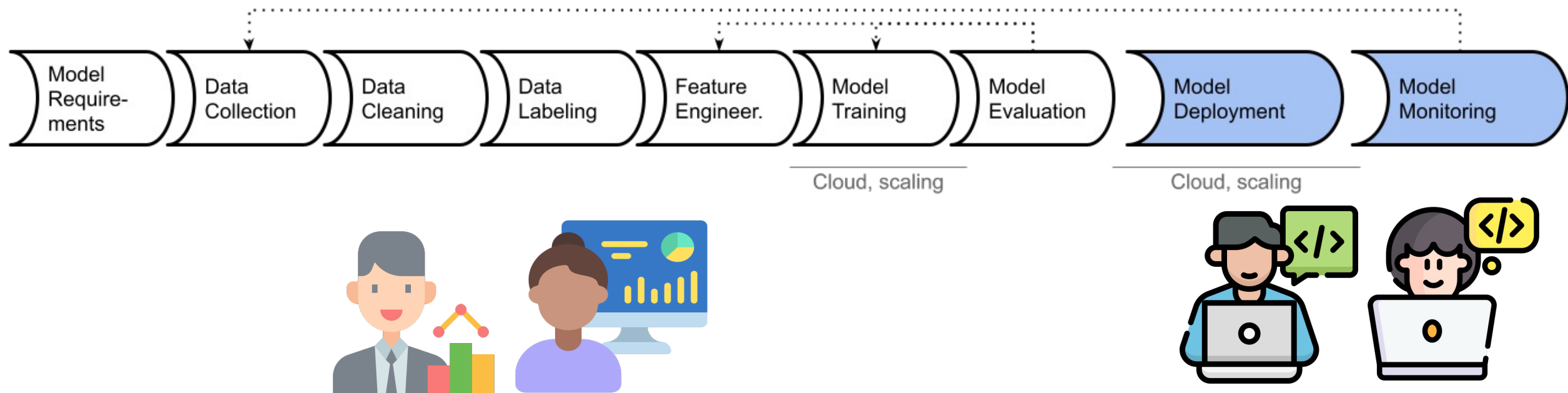
# Takeaway: Sustainability problem can get worse

- Two-fronted risk
- Requires both domain-specific knowledge & SE knowledge



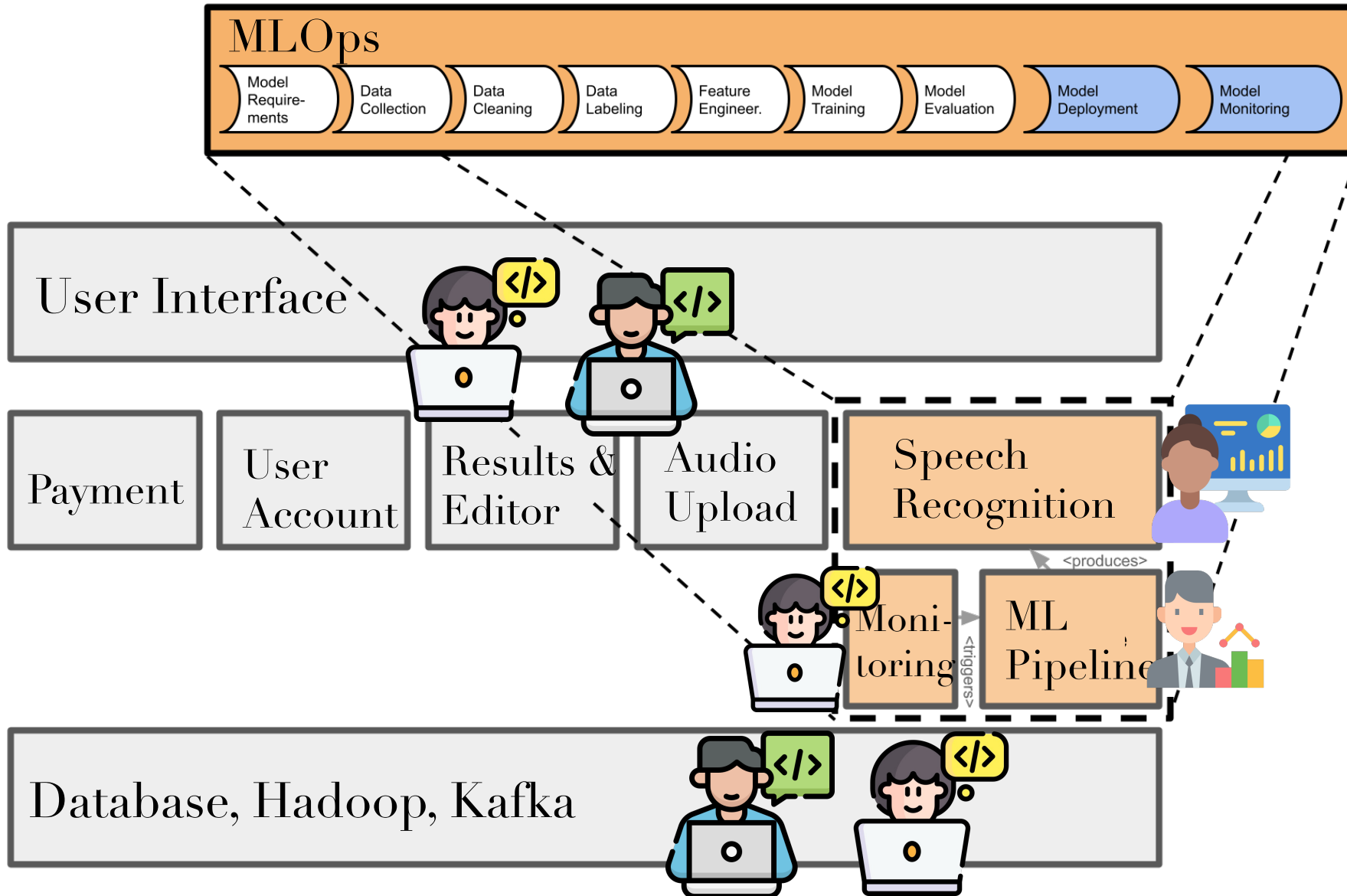
# Interdisciplinary Collaboration when building AI-based Software

- Different experts tend to focus on different stages in the machine learning lifecycle



[1] Nahar et al. *Collaboration Challenges in Building ML-Enabled Systems: Communication, Documentation, Engineering, and Process*. ICSE 2022.

# Interdisciplinary Collaboration when building AI-based Software





# Our Focus: Sustainability Challenges when Building Scientific Software in Open-Source



- Majority of development work is done by scientists
- Professional SDE may be employed to create and maintain the software

[2] Kelly et al. 2015. Scientific software development viewed as knowledge acquisition: Towards understanding the development of risk-averse scientific software. *Journal of Systems and Software*.

[3] Katherine A. Lawrence. *Walking the Tightrope: The Balancing Acts of a Large e-Research Project*. CSCW 2006.

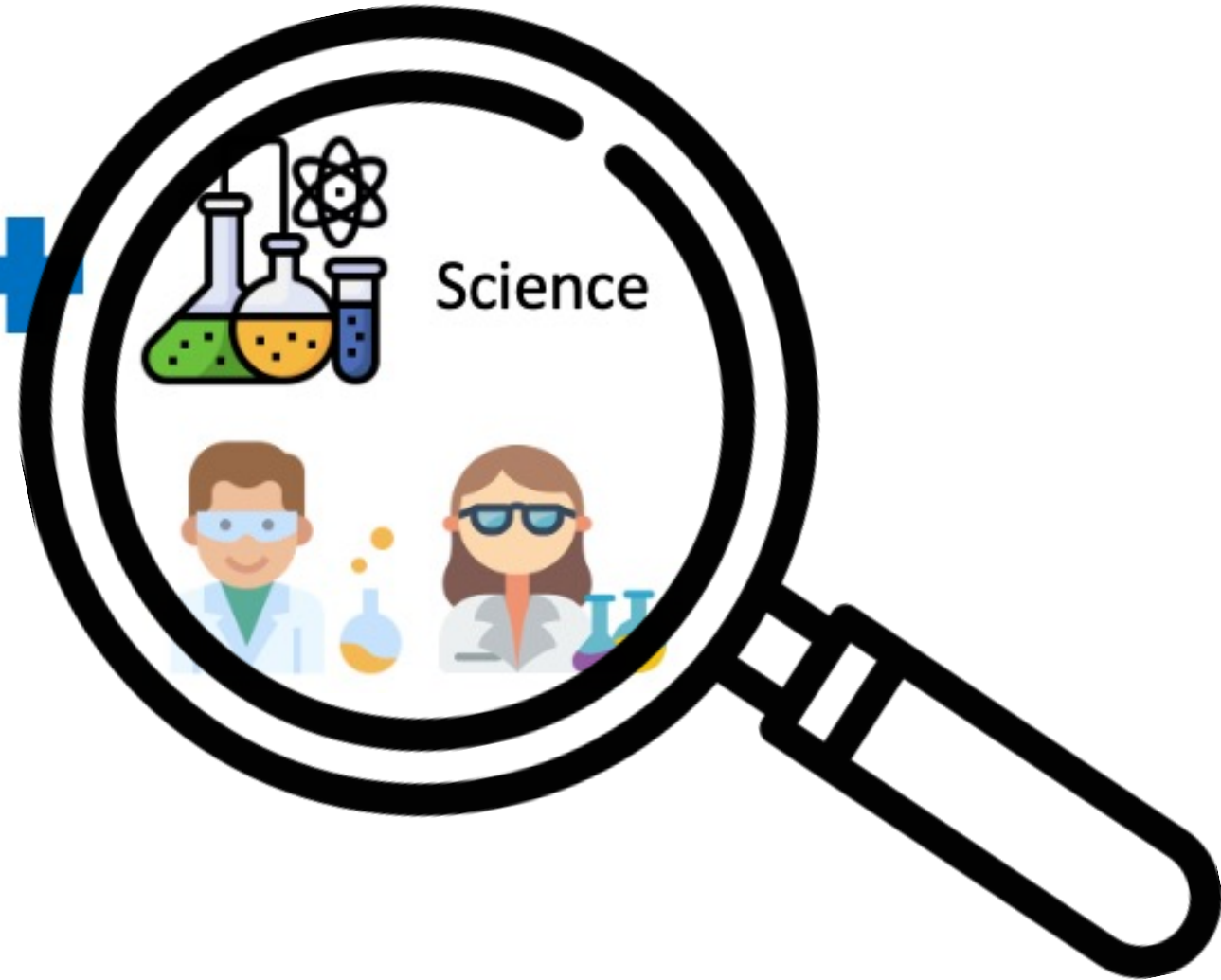


Q1

What are the major obstacles when an interdisciplinary team builds and maintains a scientific OSS?

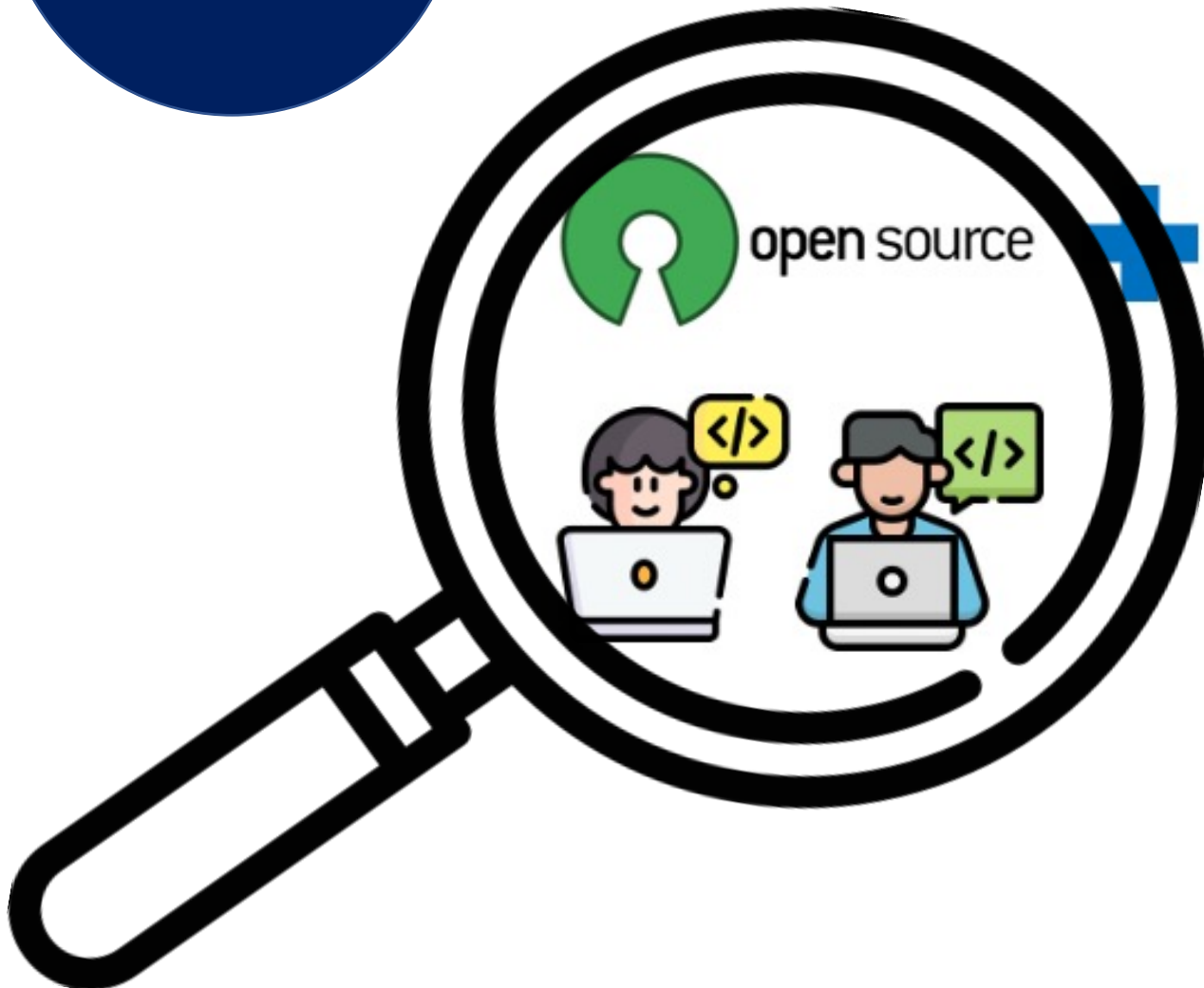


Science



Q2

What are the main factors for sustaining the scientific OSS community?



Science



# A Case Study on a scientific software in Physics domain



# A brief intro of Moonpie

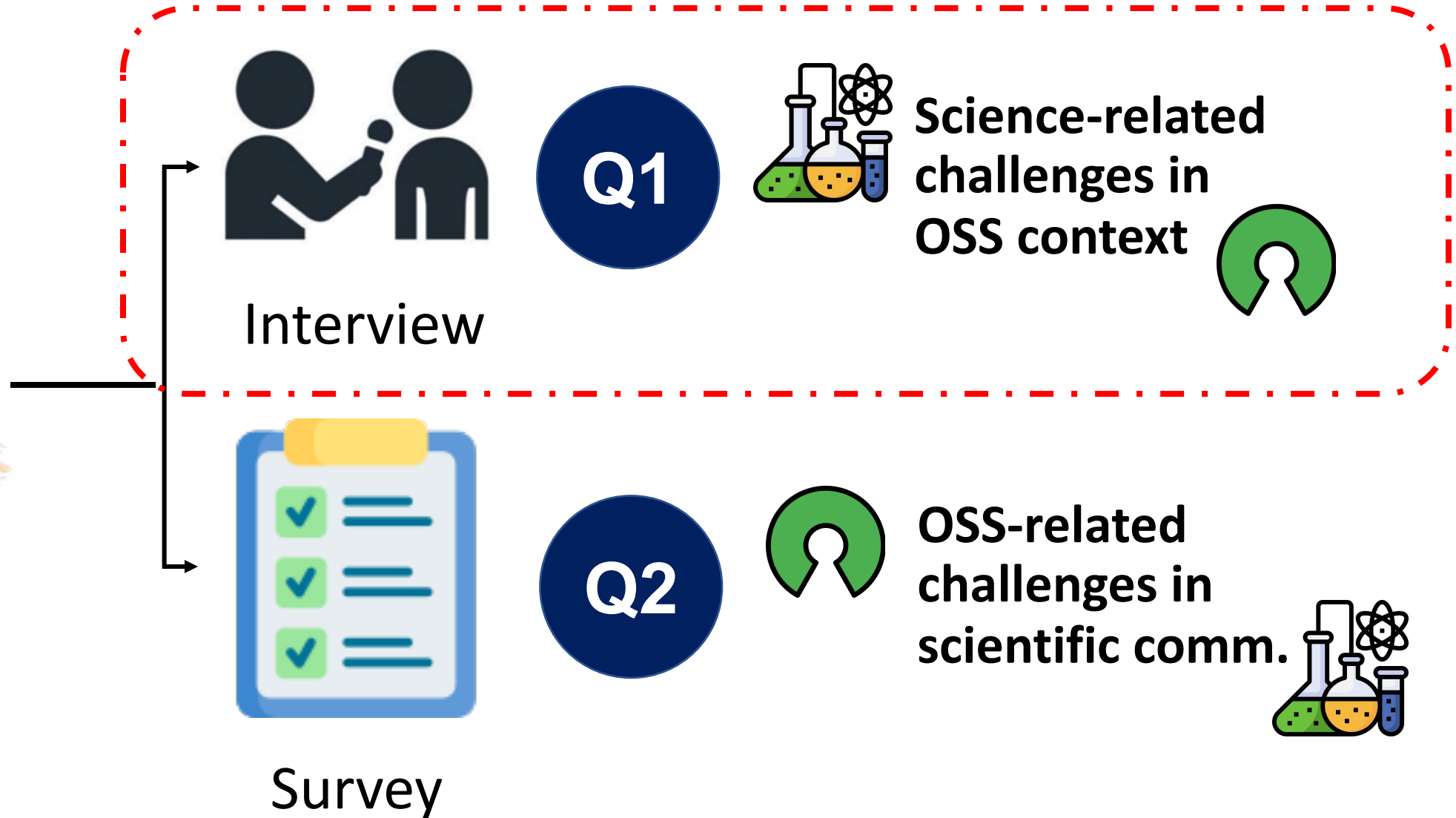
- A software ecosystem
- 1 core package & 50 interoperable packages
- Core package has 1.6K forks
- Over 10 years
- The project has been cited over 5.5K times
- 41 core contributors
- > 400 contributors in total



# A Case Study on a scientific software in Physics domain



Mining  
Repositories



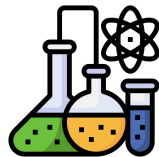
# Q1: Science-related challenges in OSS context

- Focusing on 41 core contributors
- Understanding the type of contributions
  - The code file can be divided into 2 categories:

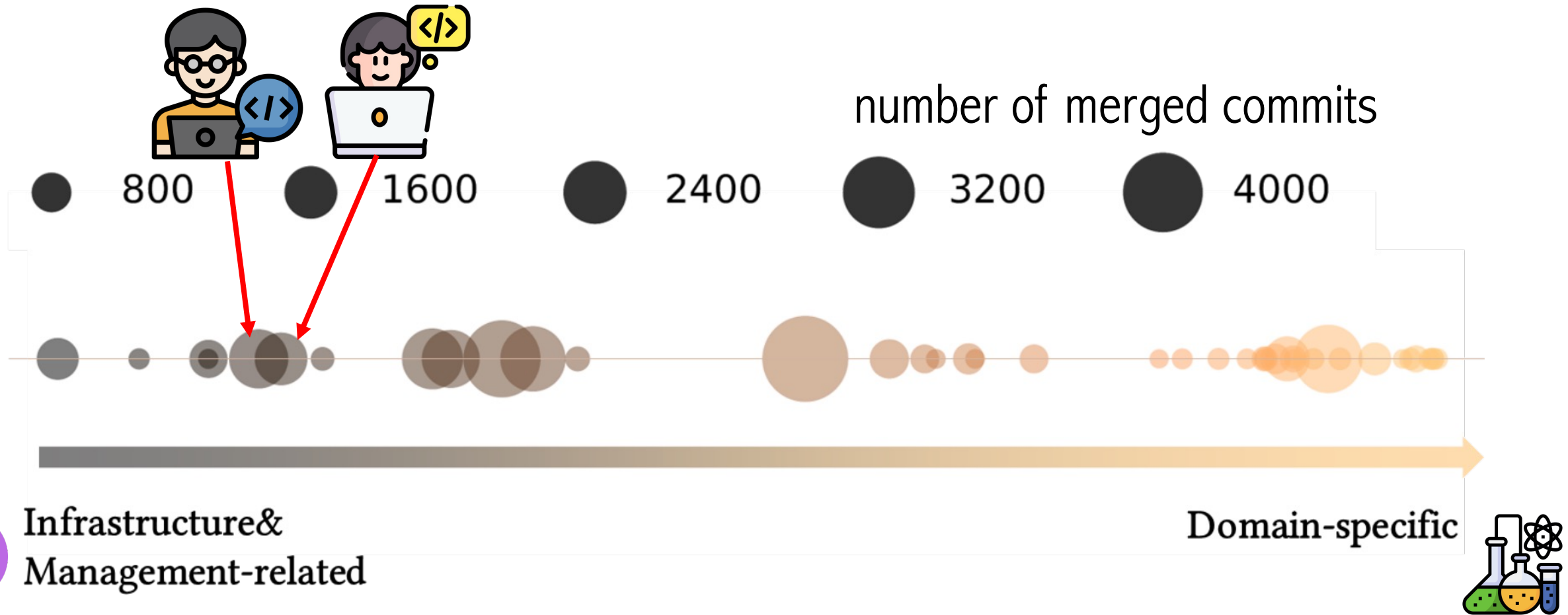
- Infrastructure



- Domain-specific

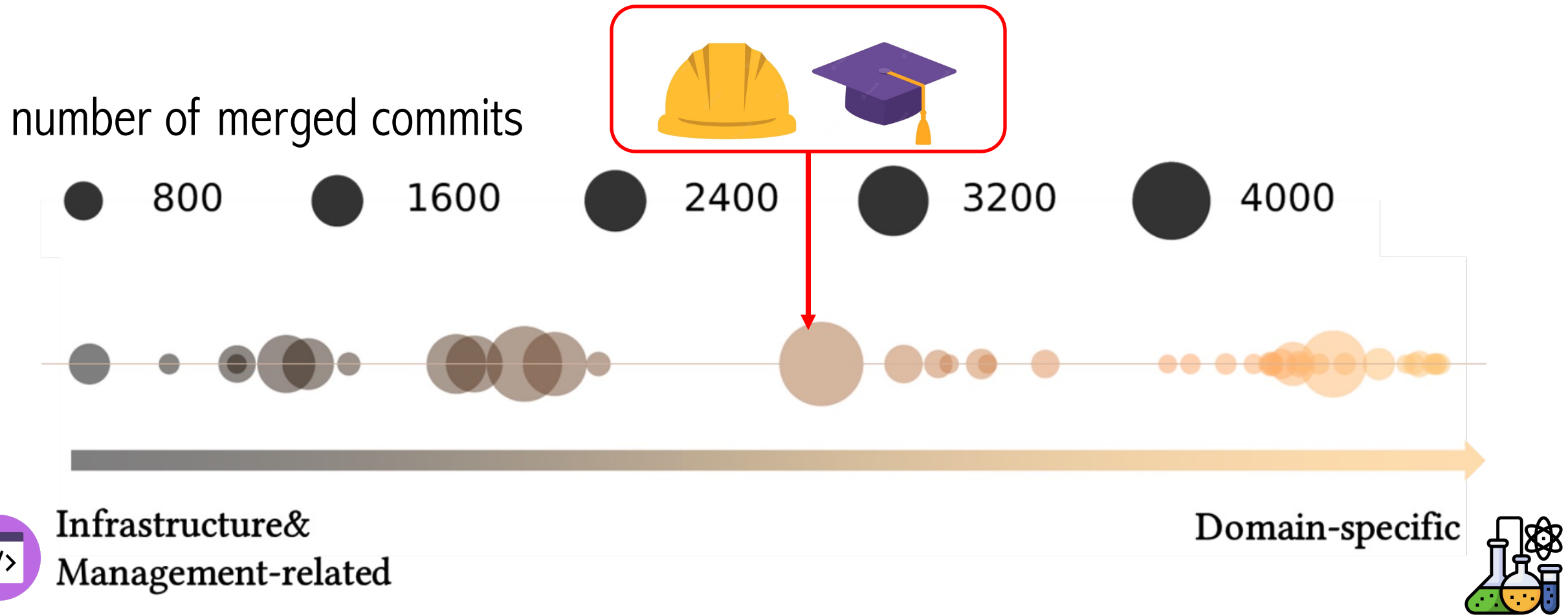


# Q1: Science-related challenges in OSS context





# Q1: Science-related challenges in OSS context

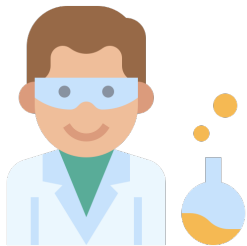


# Conflicts between different mindsets

- Incentives of making contribution
- Prioritizing the tasks



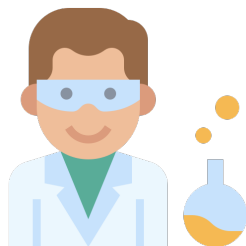
*“rigid coding standards... they need to accept the flexible nature of scientific software collaboration”*



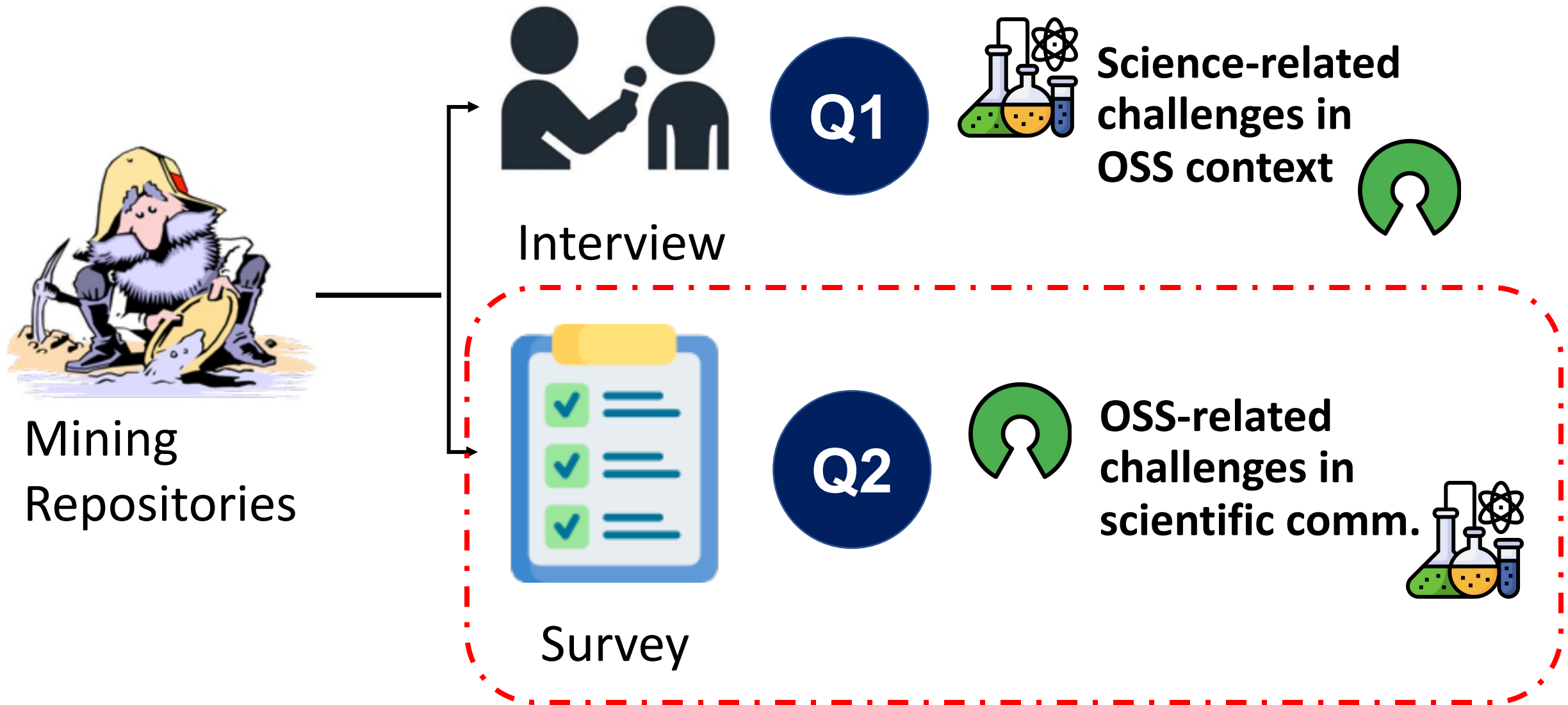
# Conflicts between different mindsets

- Perception of seniority

*“senior researchers in the decision-making position sometime ignore certain PRs because they do not see the value of the research.”*



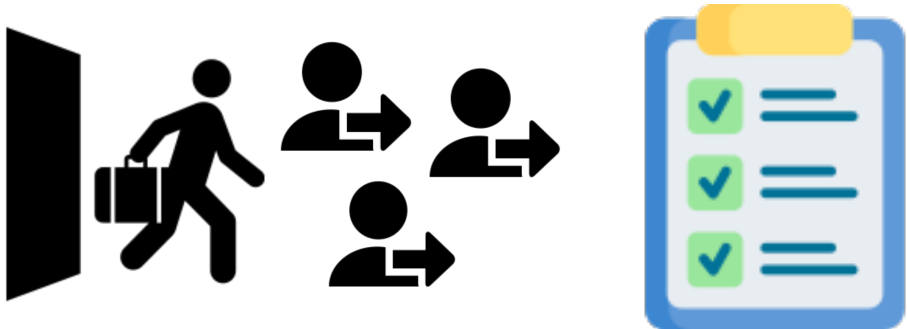
# A Case Study on a scientific software in Physics domain



# Q2: OSS-related challenges in scientific community

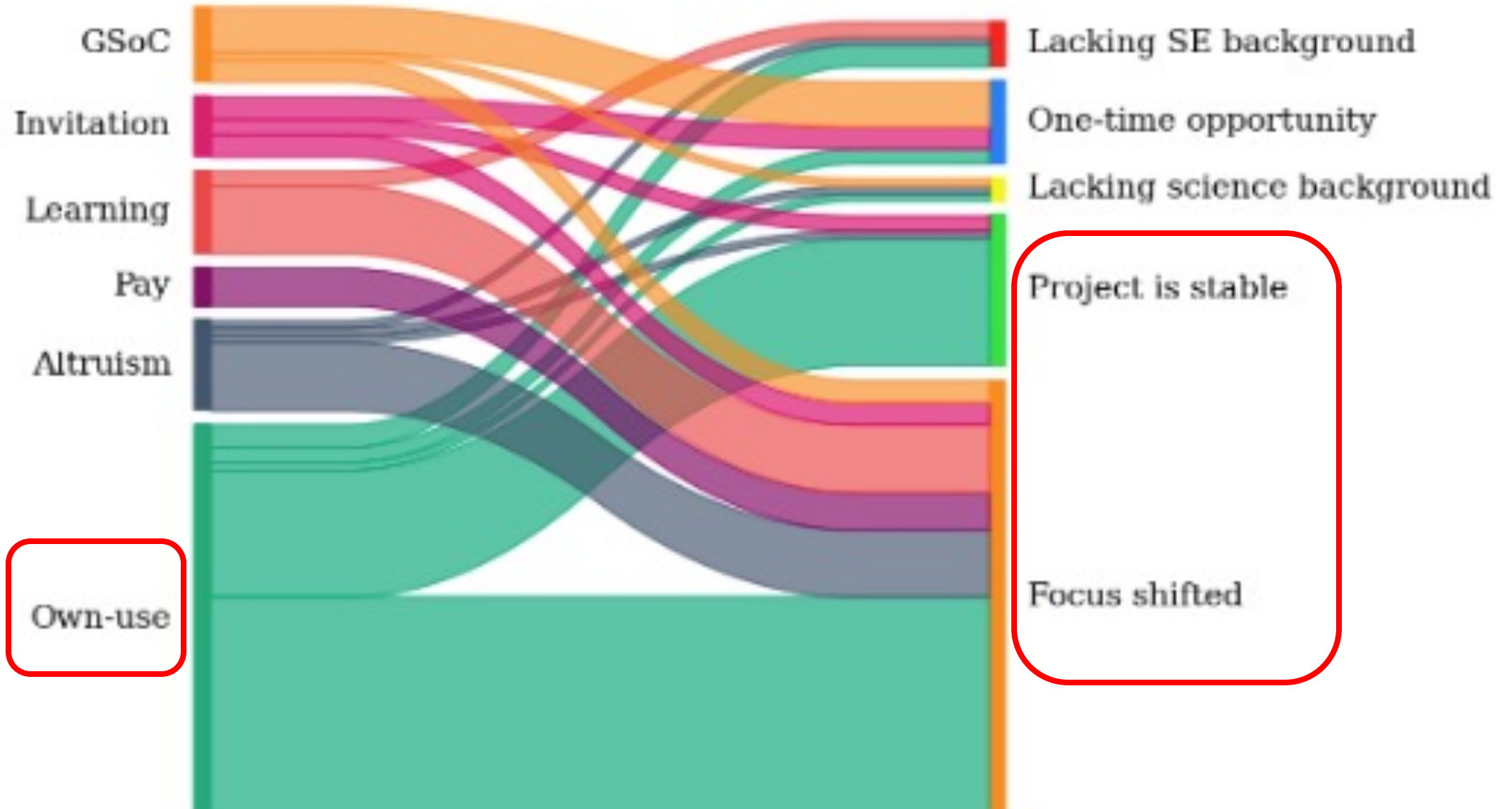
Survey questions to disengaged contributors:

1. incentive
2. reason of disengagement
3. suggestion of improving sustainability



# Incentives

# Reasons for disengagement



Own-use

Project is stable

Focus shifted



**NOT MUCH**





WHAT'S  
NEXT?

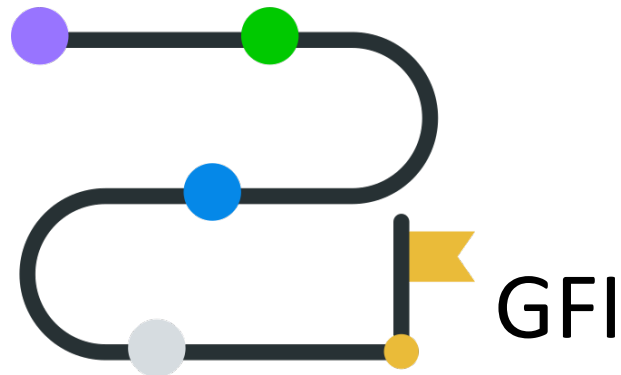
We need a  
different strategy



# We need a different strategy


## Lowering the entry barrier in science

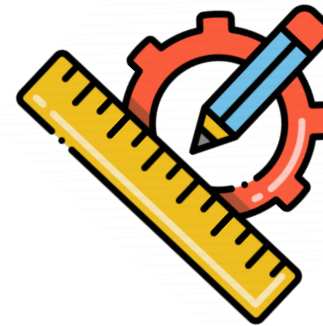
- Make both SE and science more accessible
  - Documentation for source code and scientific theory
- More guidance on Good First Issues
- Tooling support?



# We need a different strategy

## Recognizing the participation and contribution

- Citing the tool you are using 
- Acknowledging the impact of contribution
  - Quantifying the impact?
  - Identify the usage in a large scale



# Acknowledgement



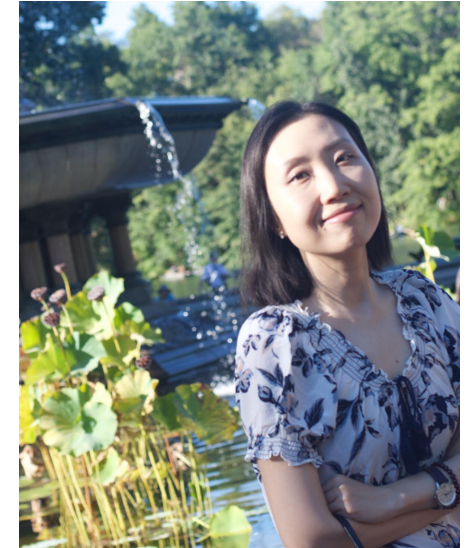
Jiayi Sun  
PhD student  
U of Toronto



Aarya Patil  
PhD student  
U of Toronto



Youhai Li  
Undergrad  
U of Toronto



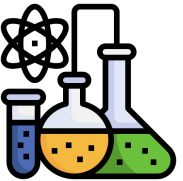
Jin Guo  
Professor  
McGill University



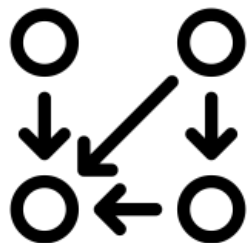
# Understanding the Sustainability Challenges for Building Open-Source Scientific Software



Recognize the unavoidable tension



Improving accessibility for both code and science



Giving credits to the contribution

**Shurui Zhou**



@shuishuiblue



UNIVERSITY OF  
**TORONTO**