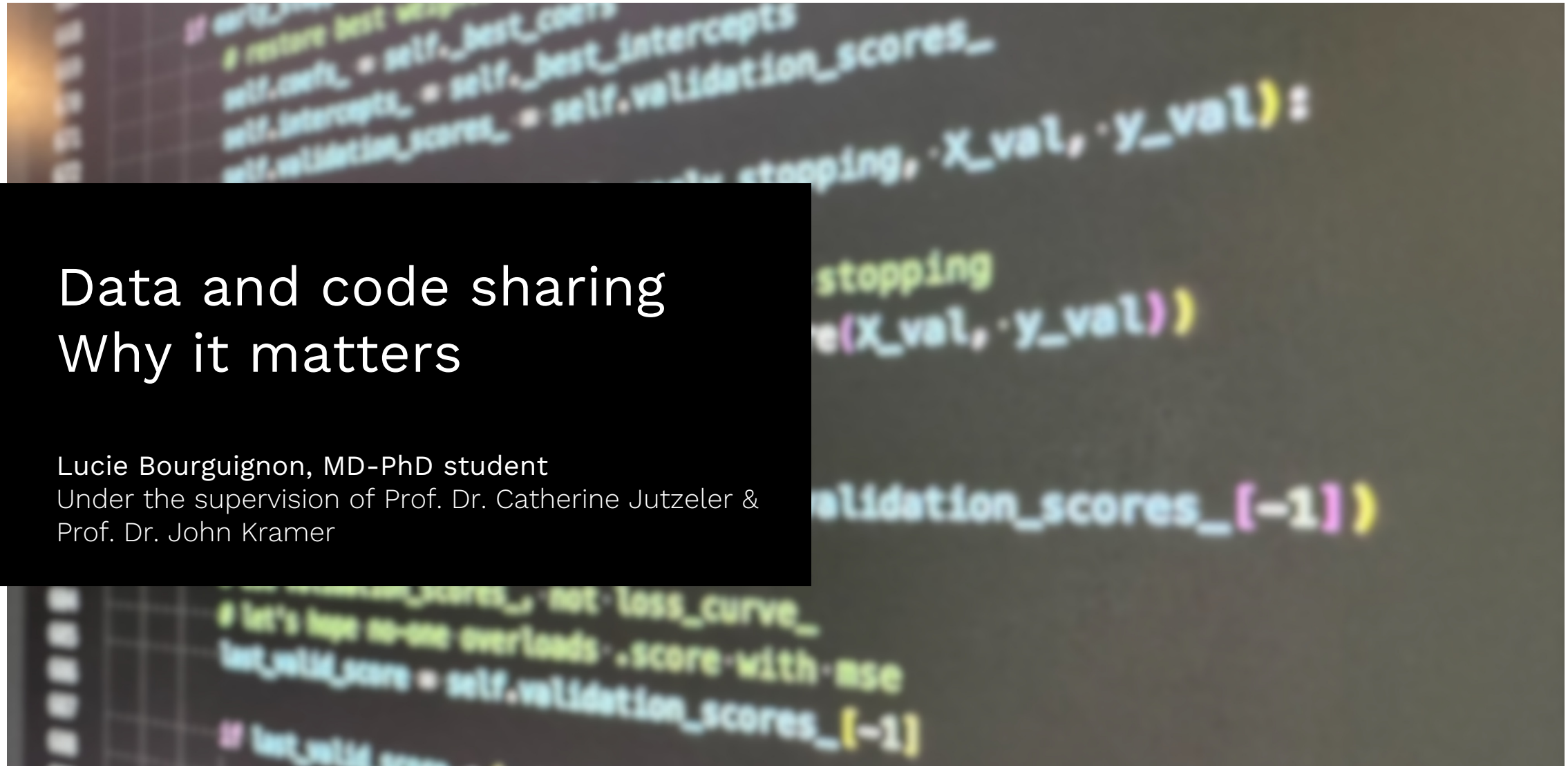
ETH zürich

Kramer Lab

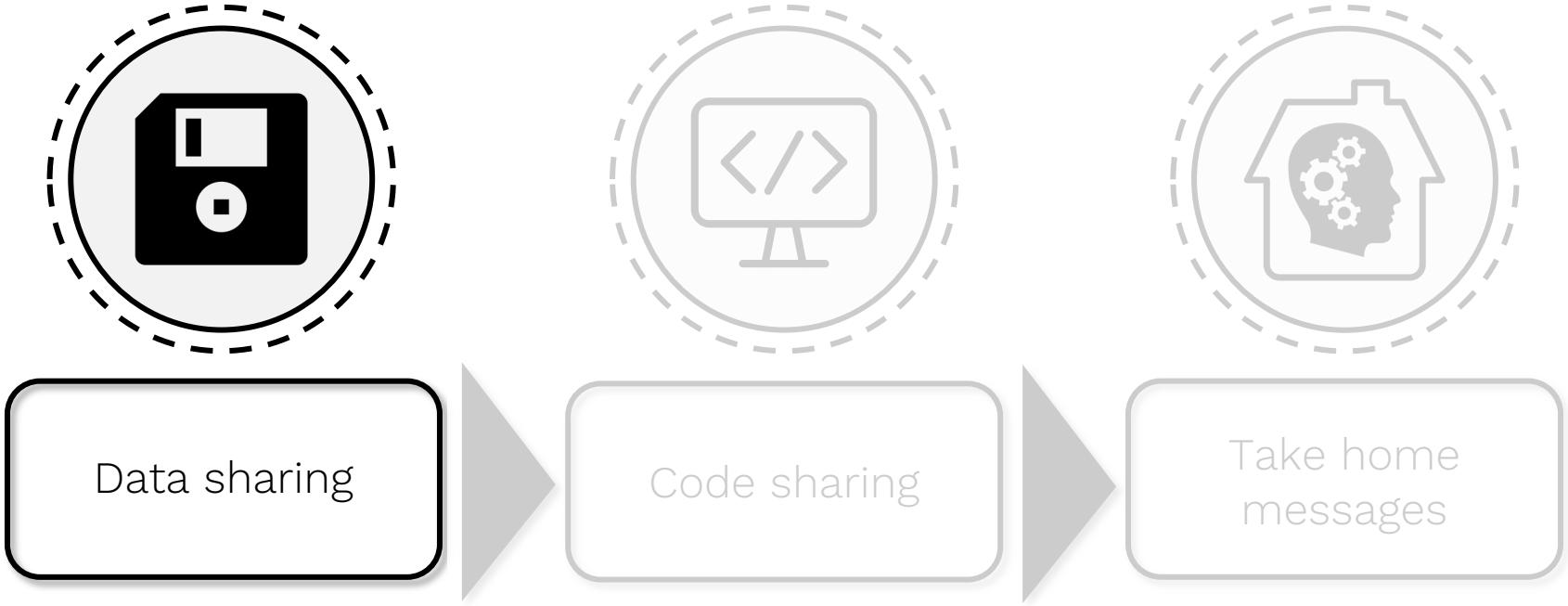
icord



Data and code sharing
Why it matters

Lucie Bourguignon, MD-PhD student
Under the supervision of Prof. Dr. Catherine Jutzeler &
Prof. Dr. John Kramer

Outline



Data sharing
Introduction

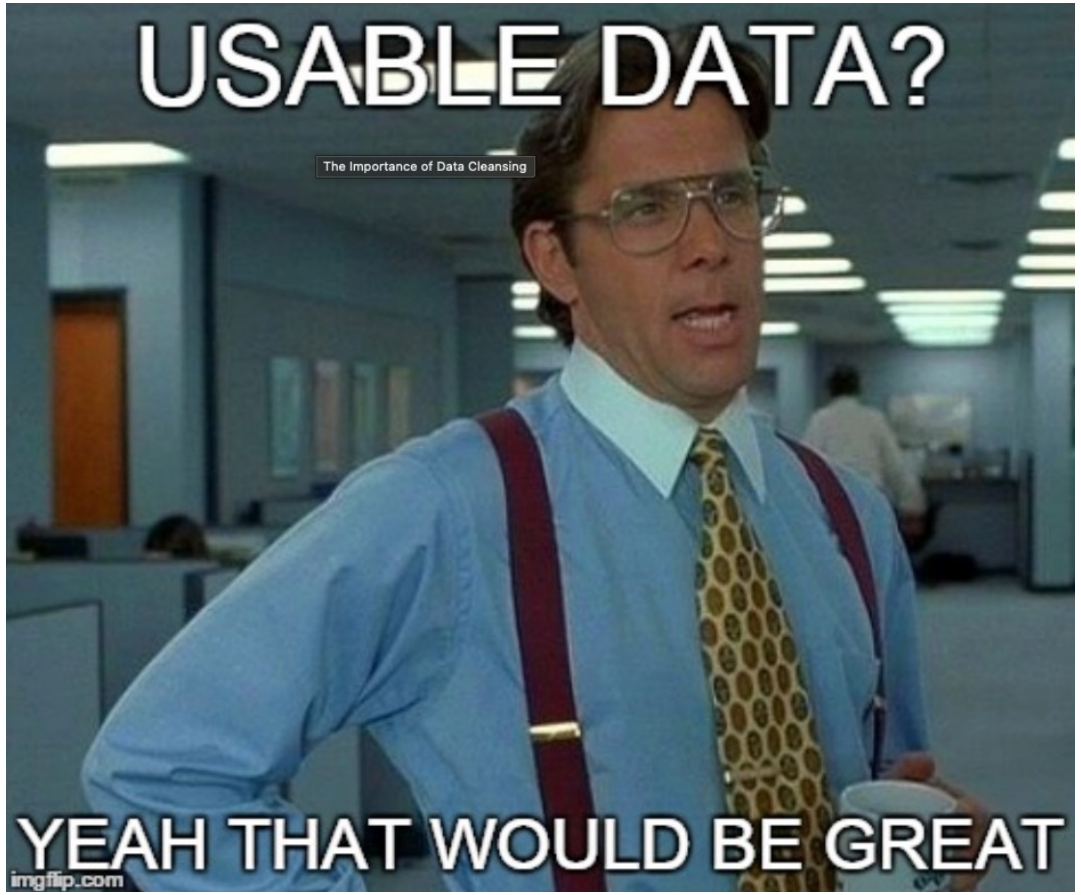
Process of making the same data resources available to multiple applications, users, or organizations, while maintaining data fidelity across all entities consuming the data. [source]

1 Have you ever used data from others?

2 Have you ever shared your data with someone?

Data sharing
Sharing locally

1. The data you collect is a property of the lab.
2. Common formatting is important (cf Jessie's lab meeting presentation).
3. Document your data.



Data sharing
Sharing broadly – FAIR principles

F indability

A ccessibility

I nteroperability

R euse of digital assets

Guideline for those wishing to enhance the reusability of their data holdings. Distinct from peer initiatives that focus on the human scholar, the FAIR Principles put specific emphasis on enhancing the ability of machines to automatically find and use the data, in addition to supporting its reuse by individuals. [source]

FAIR principles and original paper introducing FAIR Data Principles

Data sharing

Sharing broadly – Data sharing agreement

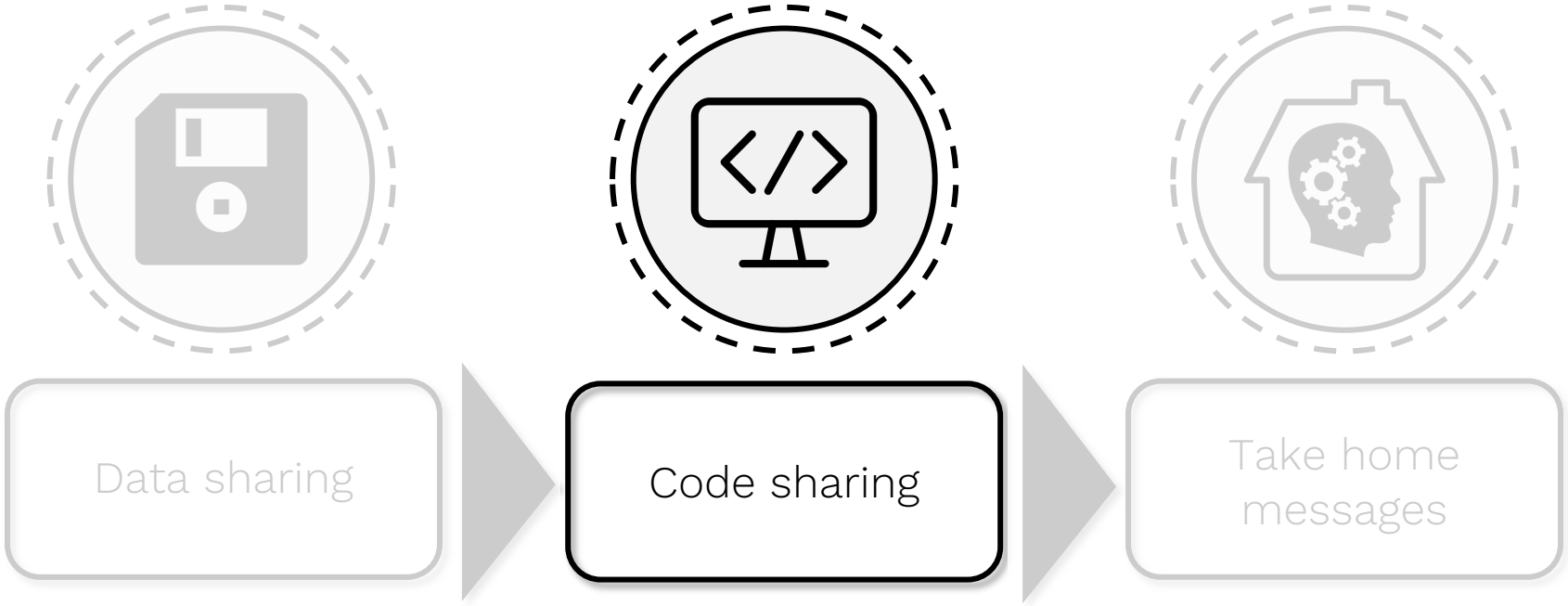
Agreement that sets out the purpose of the data sharing, cover what happens to the data at each stage, set standards and help all the parties involved in sharing to be clear about their roles and responsibilities. [source]

The further your collaborator,
the harder it gets

Data sharing
Important considerations

1. Do not share data on GitHub.
2. Do not leave patient ID in your code.
3. Do not store data locally.

Outline



Code sharing
Introduction

Range of activities, including sharing code *privately* (e.g., with your colleagues as part of internal code review) or *publicly* (e.g., as part of a journal article submission).

1. Have you ever benefited from someone sharing code with you?

2. Have you ever shared your code with someone?

Code sharing
Mentimeter

Code sharing
Benefits and barriers

- Reduce duplication effort
- Encourages transparency
- Continuity in work
- Captures work in data management

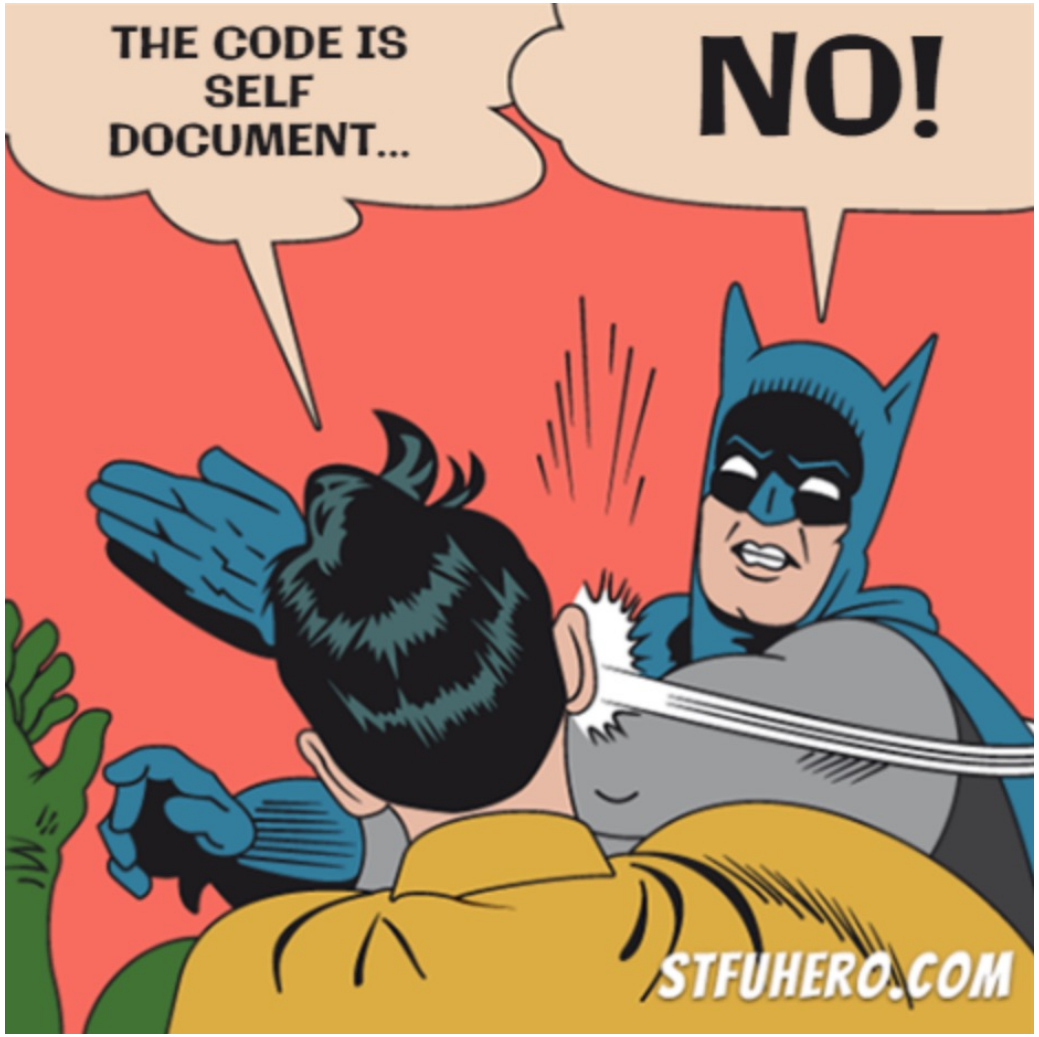
- Additional time and effort required
- Potential criticism
- Not knowing how to share
- Plagiarism

Inspired by "The benefits and barriers to code sharing as an Early Career Researcher"

Code sharing

Addressing barriers – Time & effort

1. Cleaning and documenting your code is part of the job.
2. Do it for others but also for your future self.
3. Sharing messy code is better than not sharing.



Code sharing

Addressing barriers – Potential criticism

1. Only very few people are coding experts.
2. Start with good practices as early as possible.
   see Additional resources
3. If there is a bug, it is better if someone finds it.

Code sharing

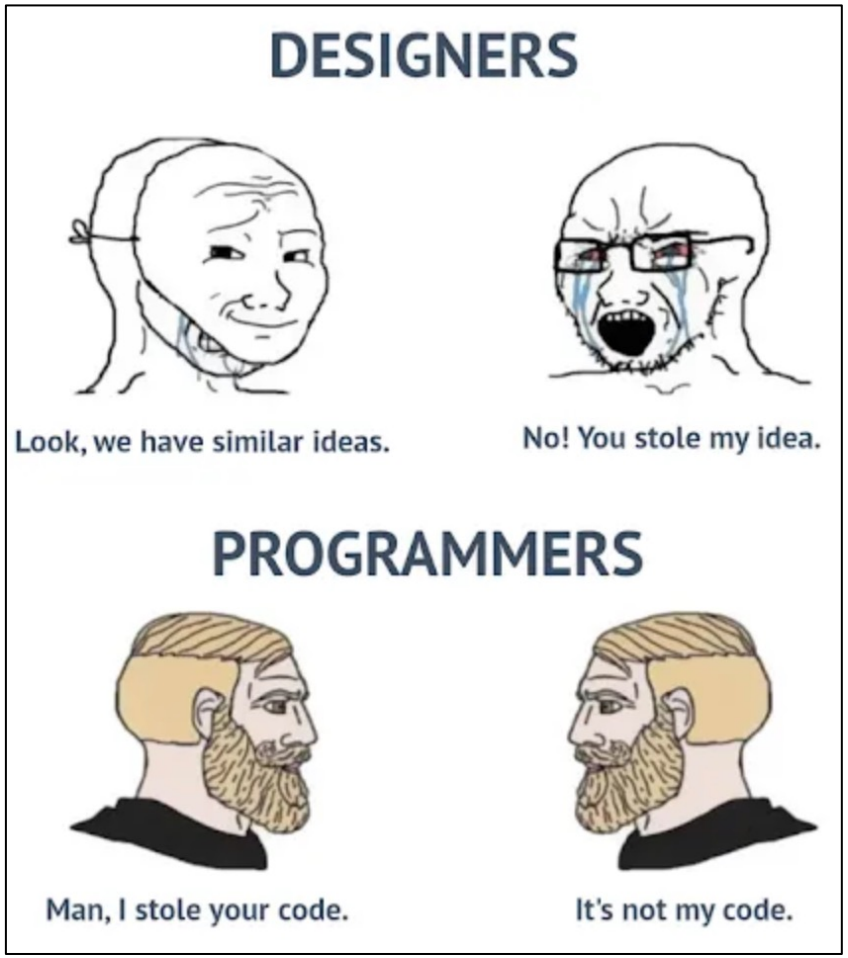
Addressing barriers – How to share

1. Use GitHub and GitHub desktop.

2. Get your code reviewed the same way you would get it for your manuscript.

Exercise break

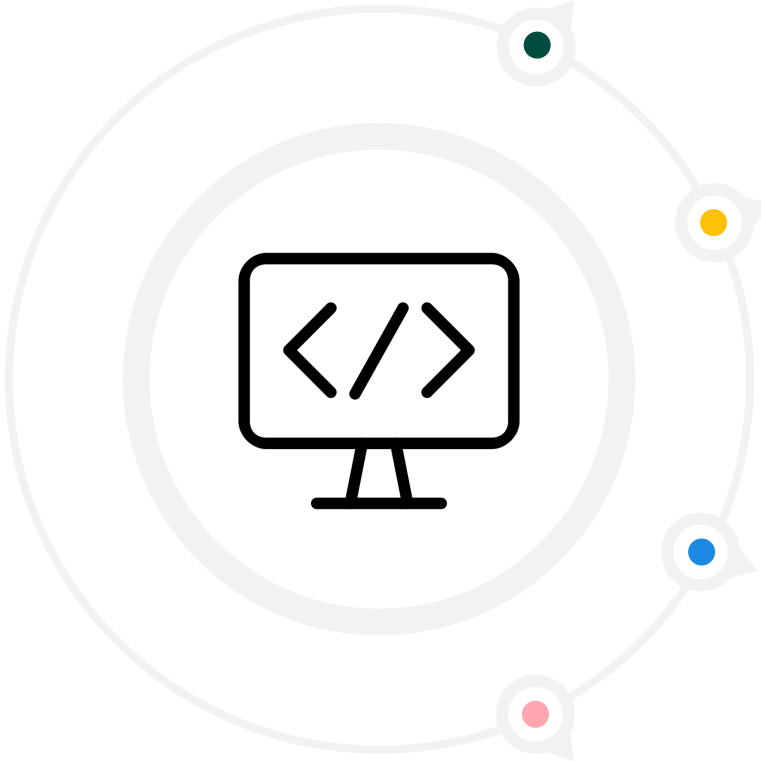
Code sharing
Addressing barriers – Plagiarism



1. Copying code from others is normal.
2. You can put a license on your code to protect it.
3. Code sharing can also be private.

Code sharing

A few coding tips



Use meaningful variable names

Structure your code

Different files, sections, functions

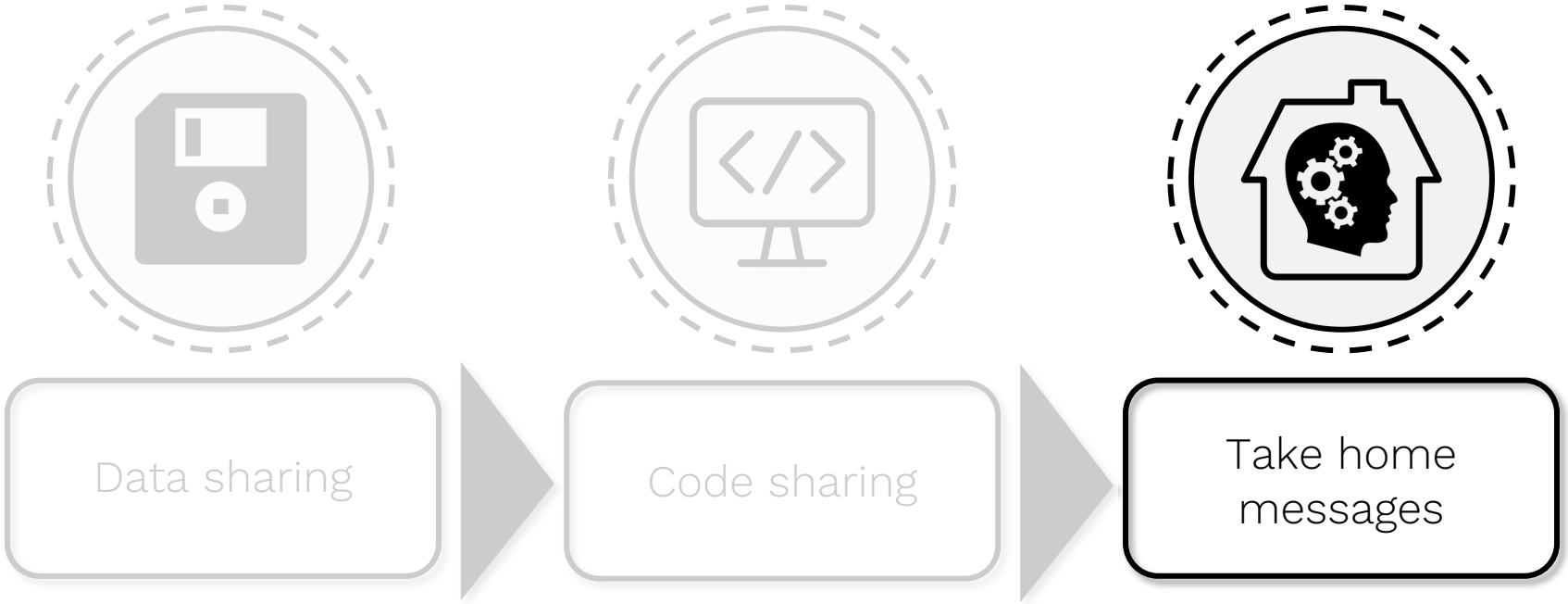
Comment your code regularly & thoroughly

Observe conventions for function descriptions

Commit on Git regularly

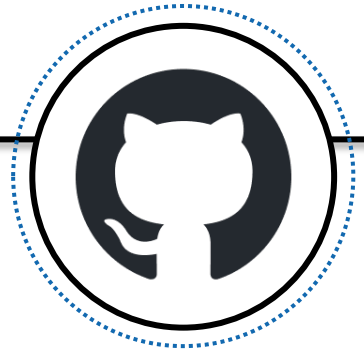
Rule of thumb: for every significant change

Outline

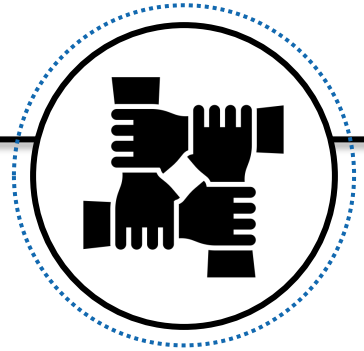


Take home messages

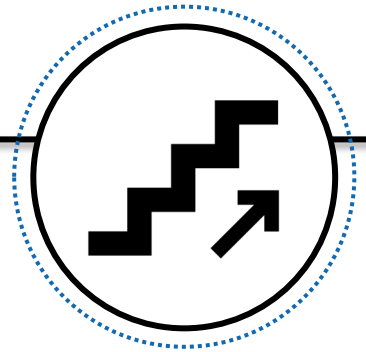
Why you should share your data & code



Contributes to making your research more open and transparent



Helps you, your colleagues and the research community



Start step by step
&
set the example

All icons used in this presentation were downloaded from thenounproject.com

Additional resources

- FAIR principles
- Original paper introducing FAIR Data Principles
- The good research code handbook
- Learning Git branching
- GitHub Skills
- GitHub Desktop