# Unit 1: Reproducibility and Replicability in Scientific Research - Instructor Notes

Part of the Learning Path: Technical skills as the bridge to reproducibility: an introduction for data librarians.

Authors and instructors:

* Agnes
* Tuulikki

Training scheduled for: Friday, 17th January 2025

Changes made: [**UNIT 1\_LearningPath\_DraftSlides\_V3\_Nov2024.pptx - ONLYOFFICE**](https://workplace.skills4eosc.eu/Products/Files/DocEditor.aspx?fileid=7848)

Update? [LearningPath\_DraftSlides\_July2024\_v02 (Units 1 and 2).pptx - ONLYOFFICE](https://workplace.skills4eosc.eu/Products/Files/DocEditor.aspx?fileid=7848)

## Notes per slide

### Title slide (slide 1)



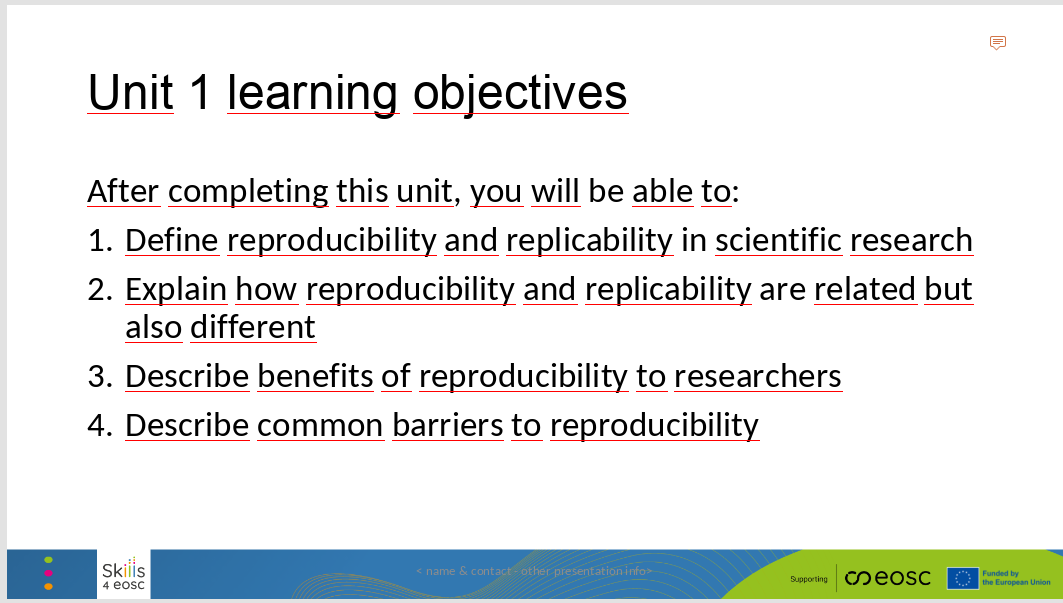
Welcome. This session will focus on reproducibility and replicability in scientific research. It is the first unit in the learning path titled, *Technical skills are the bridge to reproducible research. An introduction for data librarians*, developed for the Skills4EOSC project.

### Slide 2



The instructors for this unit are Agnes Jasinska and Tuulikki Alamettälä.

### Slide 3

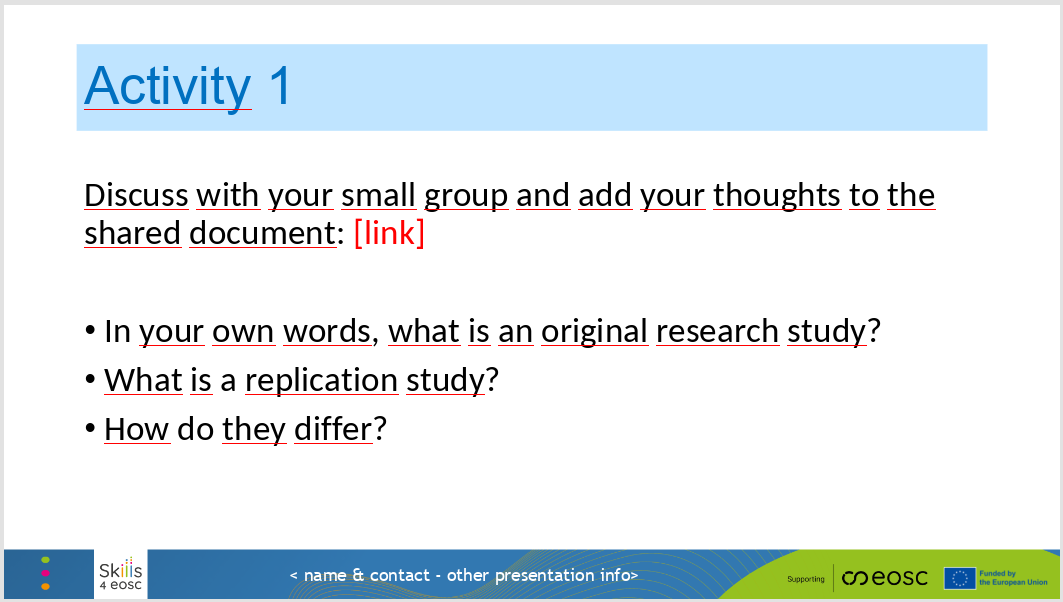


Here are our learning objectives for today’s session.

By the end of this session, you should be able to:

1. Define reproducibility and replicability in scientific research.
2. Explain how reproducibility and replicability are related but also different.
3. Describe benefits of reproducibility to researchers.
4. Describe common barriers to reproducibility.

### Slide 4



This brings us to a short group activity. We will divide you into groups and send you into breakout rooms.

Once you are with your group, here are three questions that we would like you to discuss.

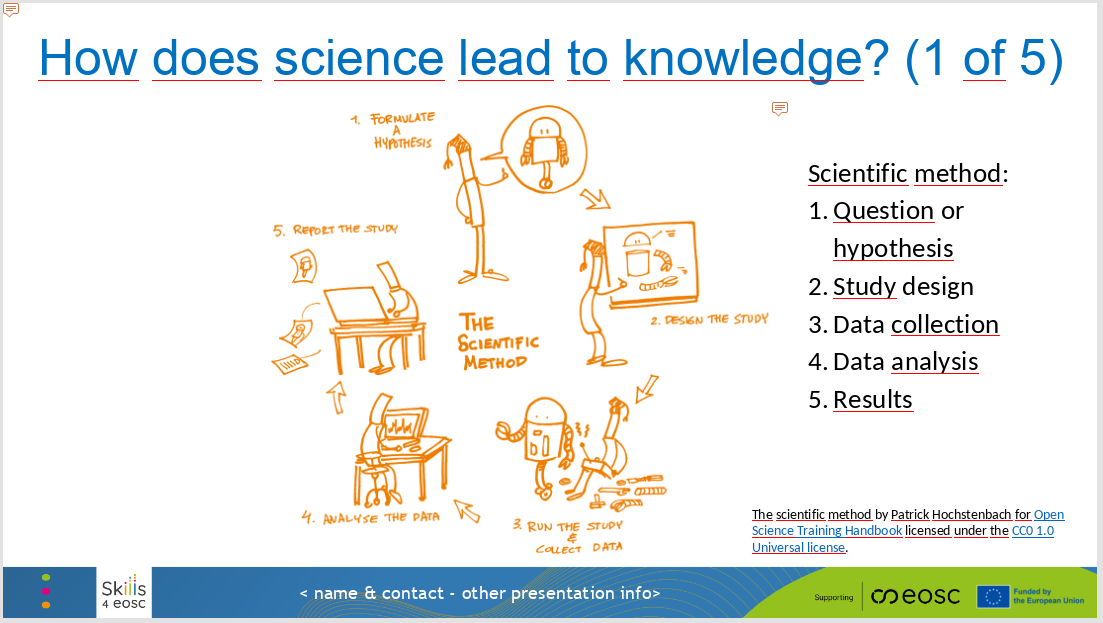
1. In your own words, what is an original research study?
2. What is a replication study?
3. How do they differ?

Please add your thoughts to the shared document. The link is in the chat.

*[ Implementation notes: In the first step, learners discuss in small groups and record their ideas in a shared document. Alternatively, each learner could reflect on their own and add ideas to the chat. In the second step, the instructor brings the whole class together to discuss.*

*The goal of this activity is twofold. First: to let the learners think carefully about the definitions and processes we will discuss, and try to come up with their own definitions, in their own words, so the terms make sense. And second: to give the instructor a sense of where the learners are in their knowledge and understanding of the topic. Do they know the definitions and have first-hand experience applying them? Do they know the definitions but the knowledge is more abstract and not tied to real-life experience of conducting research? Maybe there are some learners coming from other scientific disciplines and using a different method? ]*

### Slide 5



We just discussed the difference between an original research study and a replication study. Both play an important role in the scientific process.

But how do we go from science to knowledge? Or from a single scientific result to reliable knowledge and understanding of some phenomenon or topic?

Let’s review the **basic steps of the scientific method**, focusing on quantitative research fields.

Step one in the scientific method is to formulate a research question or hypothesis.

Step two is to design a research study to answer that question or test that hypothesis, using appropriate research methods. For instance, we may want to conduct an experiment, which means manipulating an independent variable and measuring a dependent variable to determine cause-and-effect.

Once we decide on the appropriate study design, step three is data collection and step four is data analysis.

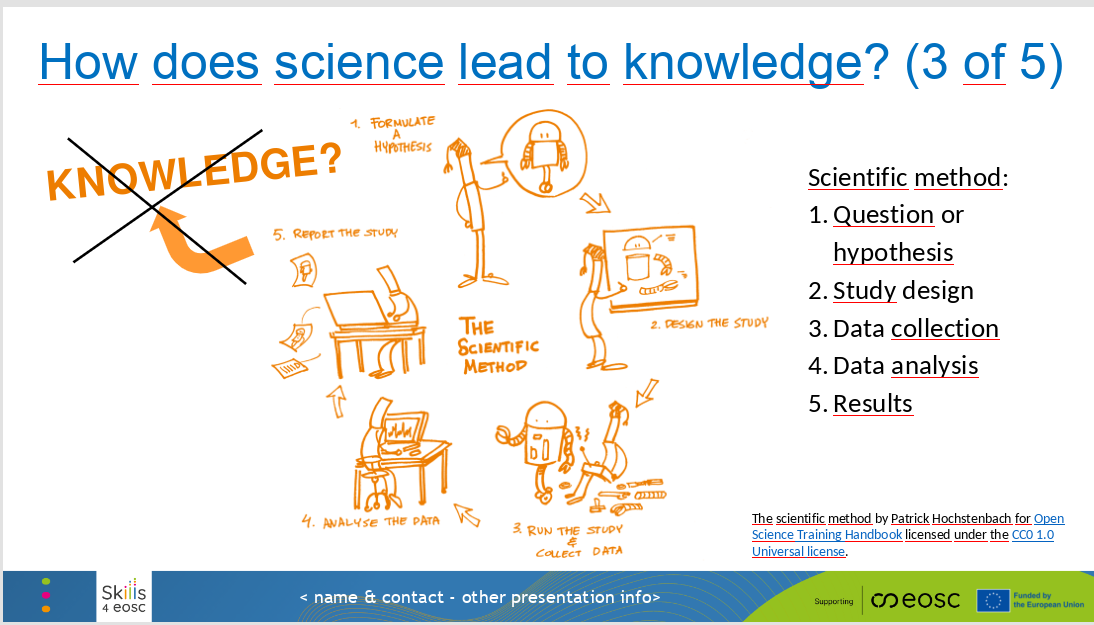
In step five, after we analyze all the data, we get our results – we get the answer to our research question or the test of our hypothesis.

### Slide 6



But is one research study and one result enough to generate knowledge?

### Slide 7

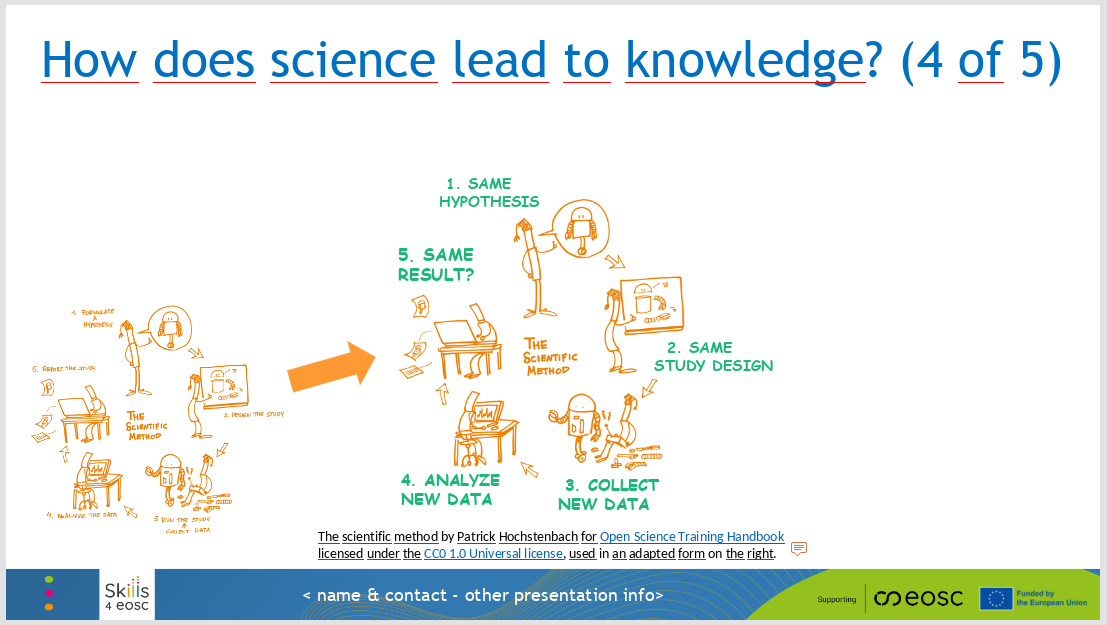


No, it is not enough.

One research study and one result can be very exciting and can give us a hint of how something works, but it is not reliable enough. We could get lucky and get a great result purely by chance. Or the phenomenon we study could be very complex and the data could be messy.

There are lots of reasons why a single research result is only the start of the scientific investigation, not the end of it.

### Slide 8

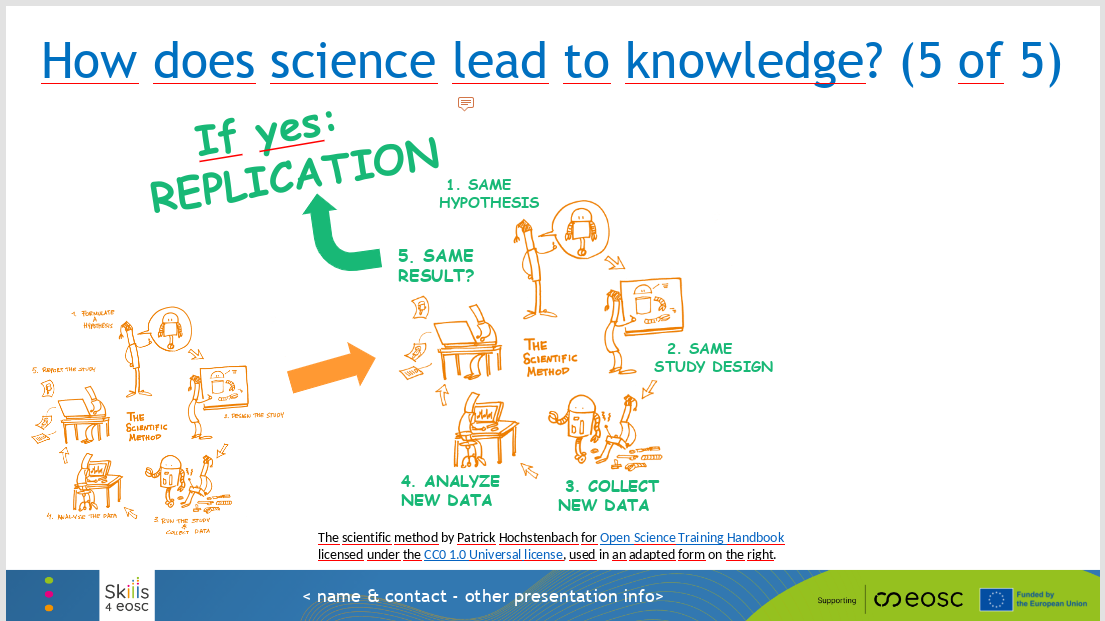


So how does science lead to knowledge? How do we make progress toward an accurate and reliable scientific understanding of some topic?

Think back to the difference between the original study and a replication study.

We iterate. We go through the steps of the scientific method once, to get our original result. And then we repeat the process and we go through the steps again to check the result.

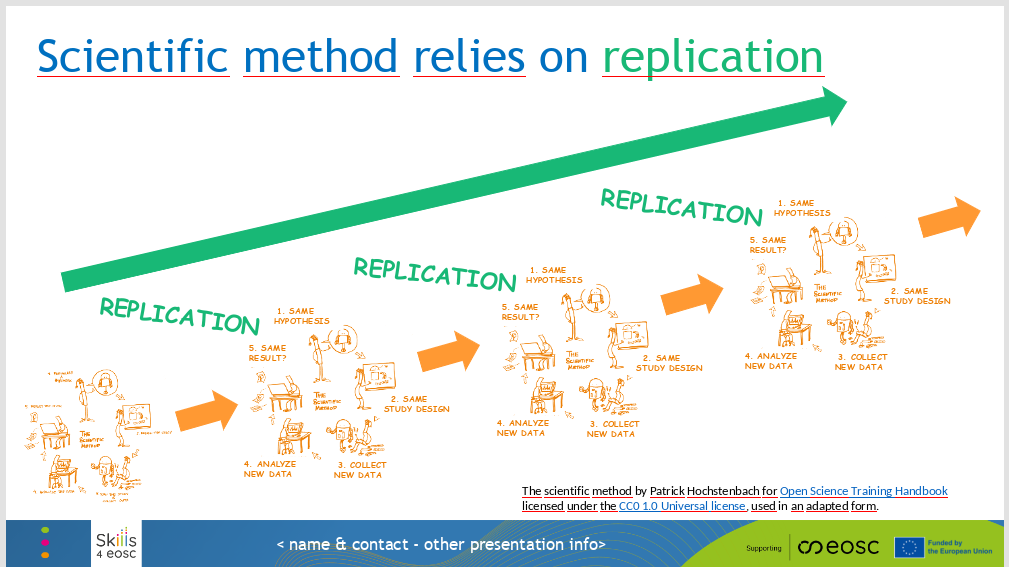
### Slide 9



Did we get the same result the second time?

If so, we have a **replication** – a second research study that produces the same result as the original research study. In other words, we got the same result twice, and now we can have a little more confidence that the result is accurate and reliable.

### Slide 10



In conclusion, **scientific method relies on replication** to give us confidence in the results.

A single research study is not enough. We want one or several carefully designed and conducted replication studies to show the same result as the first stud - before we add that result to our body of knowledge on the topic.

### Slide 11



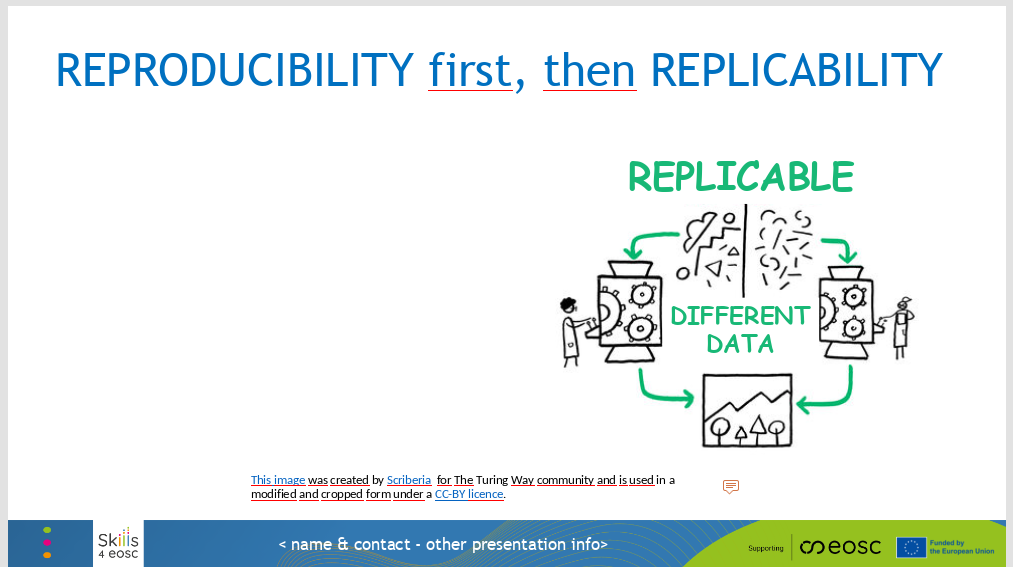
In practice, this means that **science is iterative and cumulative.**

One-shot result is not reliable scientific knowledge.

We answer the question in one dataset and report the result, then we conduct the same study using the same methods in new dataset and report that result as well for comparison. Did the result hold? What about in the same data but using new analyses? What about new data and new analyses? Do we still get a similar result?

The more we replicate a specific finding, the more reliable the knowledge and the more confidently we can build upon it. **Replication is a key step in the scientific process!**

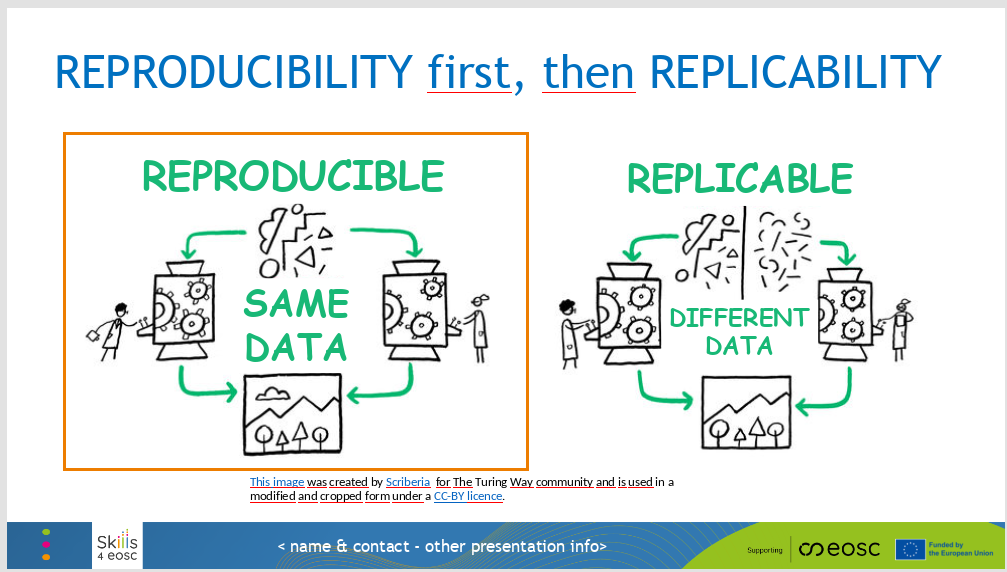
### Slide 12



So far, we focused on replication and replicability. If the original study is conducted in one dataset, the **replication study** needs to be conducted in new, different data.

In scientific research, **replicability** means that the same research result can be obtained in a new study and in new data. We can say that a research result is replicable or highly replicable, and that’s a very good thing.

### Slide 13



What about reproducibility?

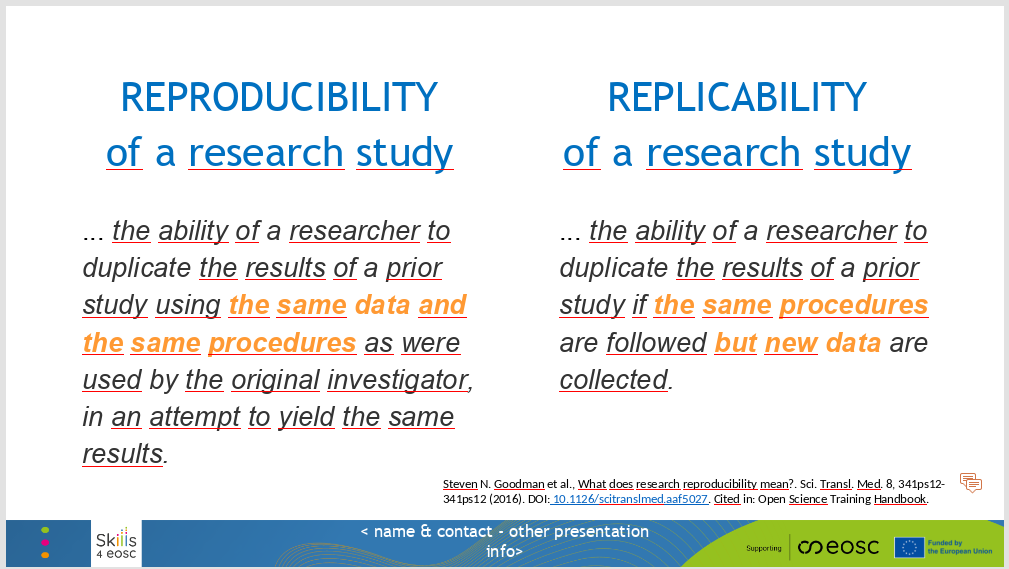
Reproducibility comes before replicability. It is even more basic than replicability.

In scientific research, **reproducibility** means that the same research result can be obtained in the same data and using the same methods and analyses.

So essentially, we are repeating the same exact study the second time – same data, same methods, same analyses. Can we get the same result? If so, then we have reproducibility.

You can think of reproducibility as checking your solution to an exam problem. If you do the problem over exactly the same way, do you get the same solution?

### Slide 14



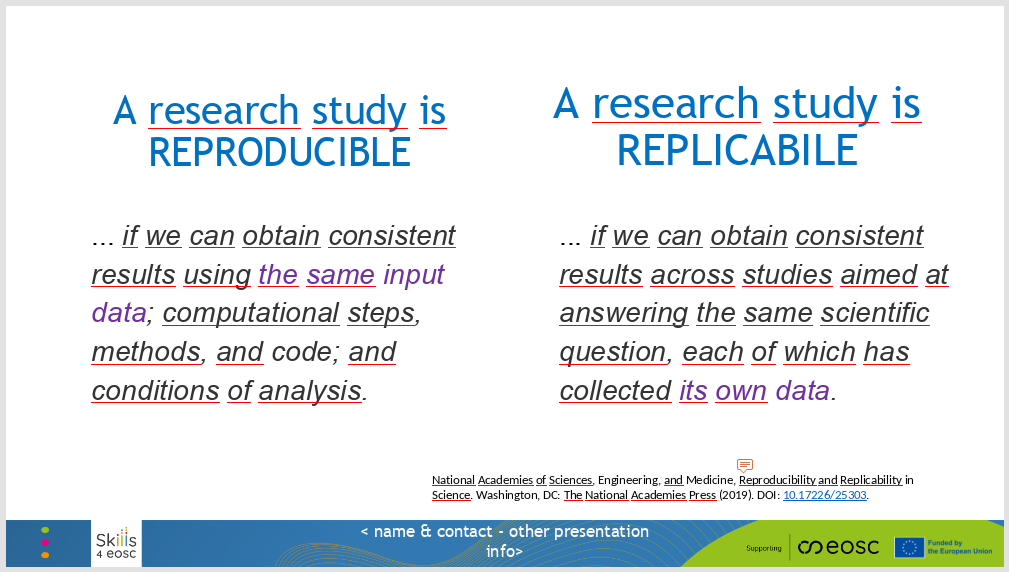
Here are the two definitions side by side.

**Reproducibility** of a research study means that a researcher can duplicate the results of a prior study **using the same data** and the same procedures as were used by the original investigator, and obtain the same results.

**Replicability** of a research study means that a researcher can duplicate the results of a prior study using the same procedures but **in new data** – and still obtain the same results.

Again, reproducibility comes first. **Replication relies on reproducibility**. If we cannot reproduce a result in the same data, how could we get the same result in new data?

### Slide 15

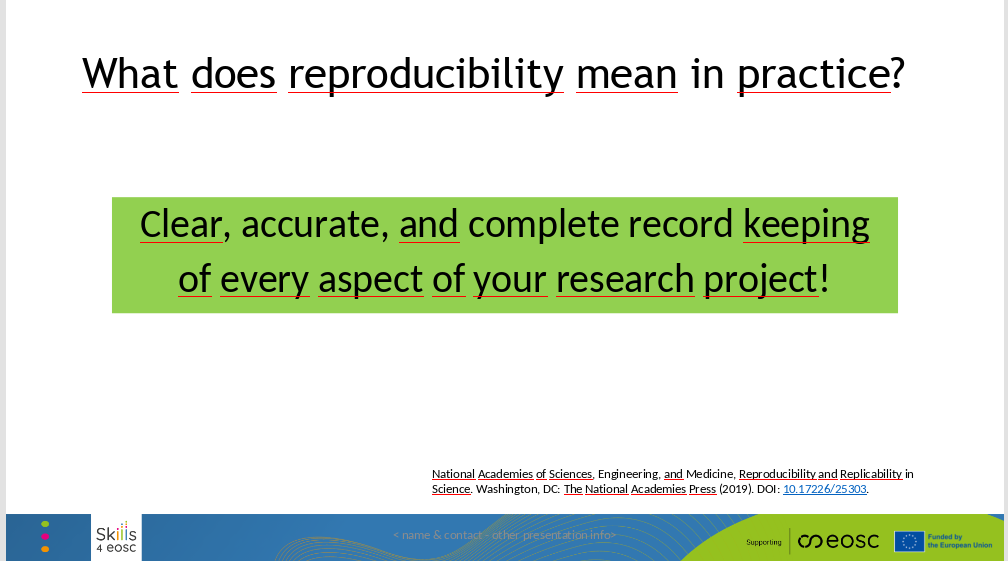


Here is another version of the two definitions, again side by side, so we can easily compare them.

**A research study is reproducible**... if we can obtain consistent results using the same input data, computational steps, methods, code, and conditions of analysis.

**A research study is replicable**... if we can obtain consistent results across studies aimed at answering the same scientific questions, but each study has collected its own data.

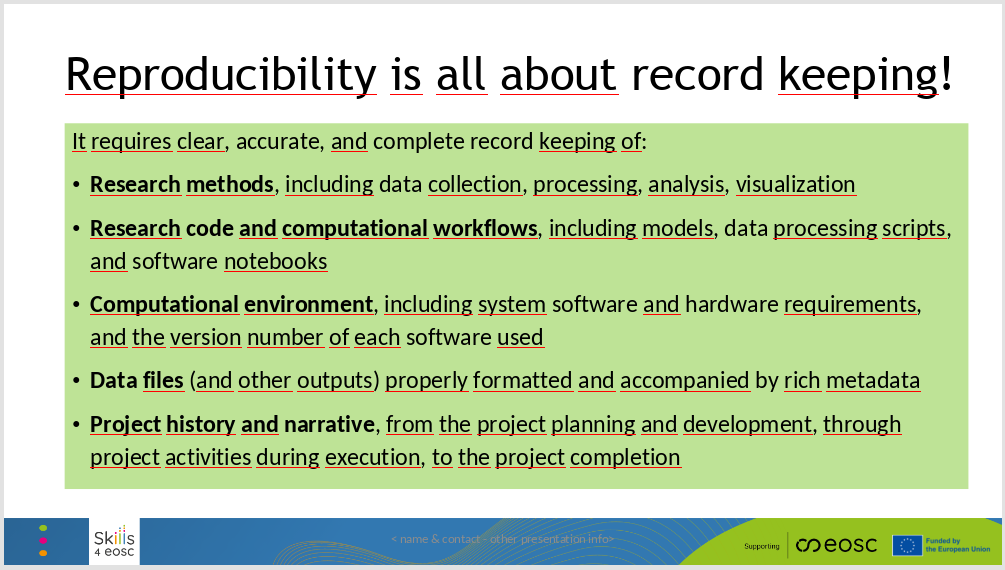
### Slide 16



What does reproducibility mean in practice?

It means **clear, accurate, and complete record keeping of every aspect of your research project**!

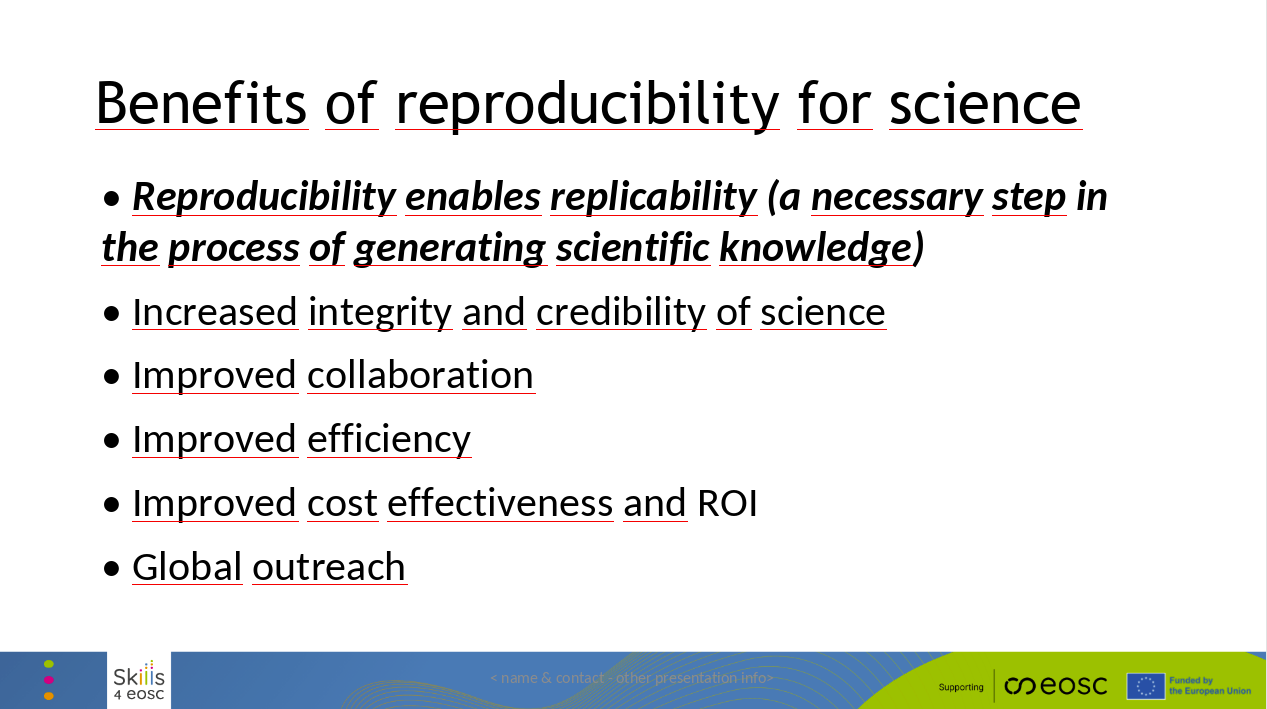
### Slide 17



More specifically, reproducibility requires clear, accurate, and complete record keeping of all of the following:

* **Research methods**, including data collection, processing, analysis, and visualization.
* **Research code and computational workflows**, including models, data processing scripts, and software notebooks.
* **Computational environment**, including system software and hardware requirements, and the version number of each software used.
* **Data files** (and other outputs) properly formatted and accompanied by rich metadata.
* **Project history and narrative**, from the project planning and development, through project activities during execution, to the project completion.

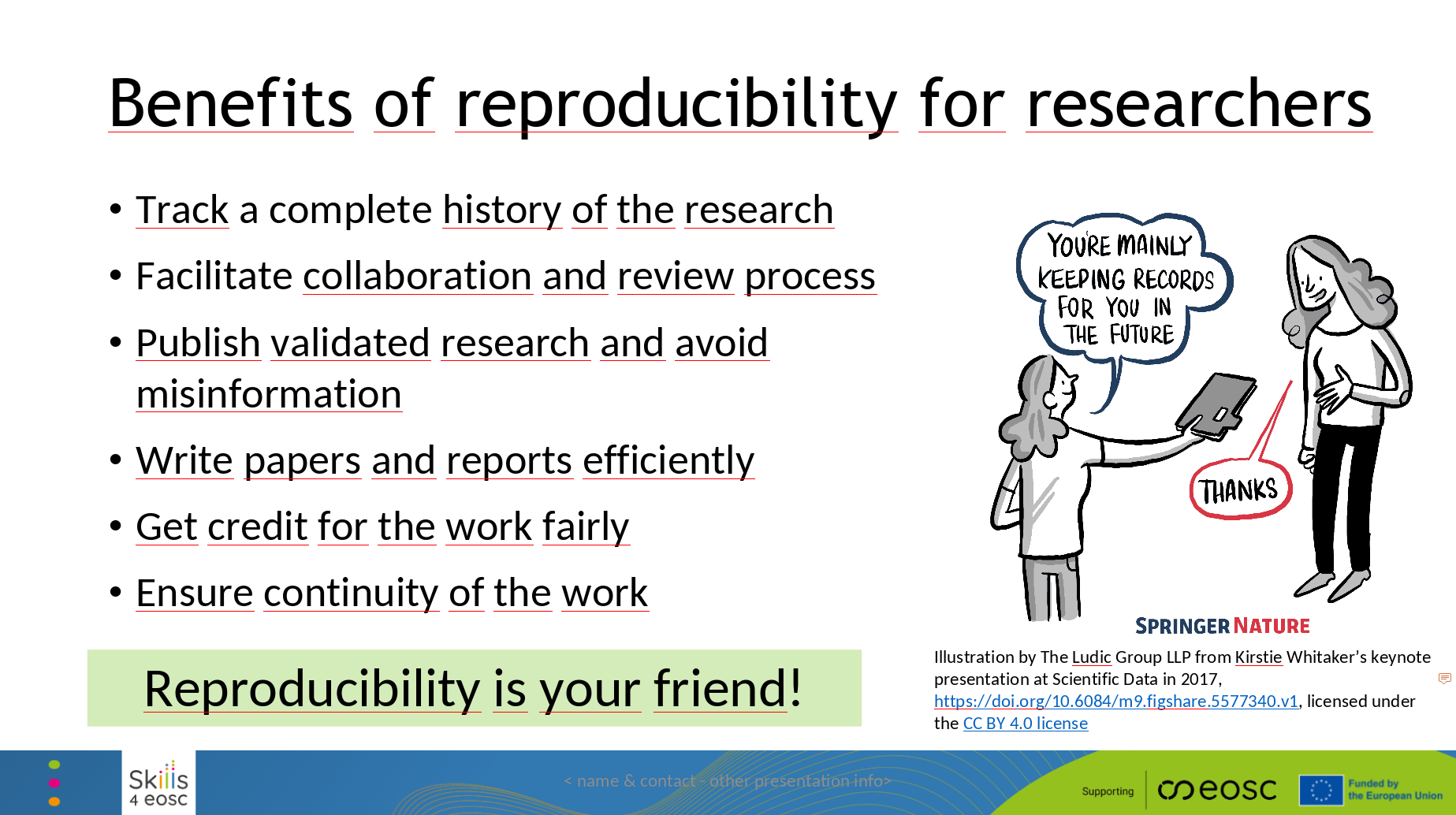
### Slide 18



Let’s talk about in a bit more detail why reproducibility is so important in science. This slide shows the many ways it benefits science.

* When research is reproducible, others can repeat the work to check if the results hold up. This is essential for building reliable knowledge and trust in science – as we have already discussed.
* Sharing the data and methods openly allows other researchers to review the work more closely. This improves the quality of research, reduces errors, and makes the whole process more trustworthy.
* It also makes it easier for researchers to collaborate, even across disciplines or countries, giving possibility to new ideas and discoveries.
* Reproducibility makes the research process more efficient. Instead of repeating the same work, researchers can use existing data to explore new questions, which saves time, effort, and money.
* Sharing research outputs helps spread knowledge faster, which drives innovation and makes research investments more impactful.
* When scientists share their work openly, it’s easier for researchers around the world to access and build on it. This helps tackle big challenges together, like climate change or public health crises.

### Slide 19

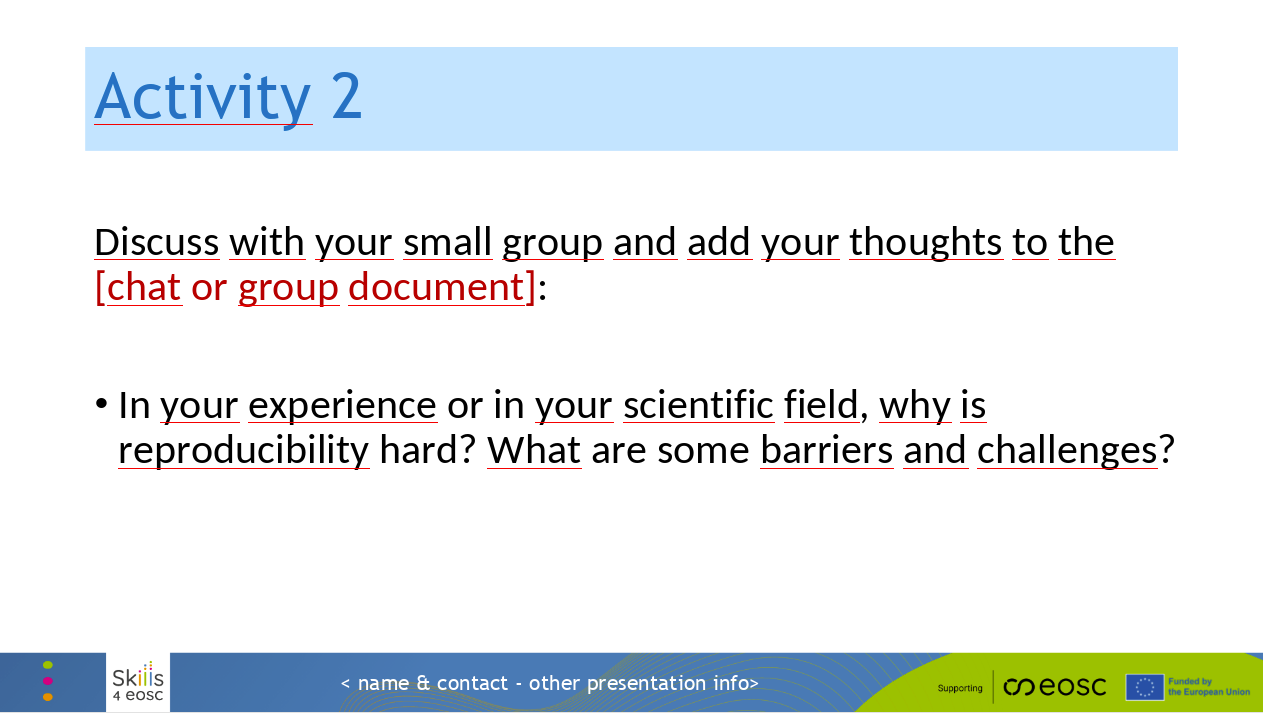


Reproducibility doesn’t just improve science—it also makes life easier for researchers. Here is how:

* Keeping a full record of the data, methods, and tools ensures that the work is sustainable, reusable, and fairly cited.
* Reproducible workflows make it easier for collaborators and reviewers to understand, test, and improve the work.
* Working reproducibly helps catch errors, avoids spreading misinformation, and ensures that the results are accurate and reliable.
* Well-documented projects make it easier to create papers, theses, and reports as everything is organized, up-to-date, and ready to use.
* Sharing reproducible work makes it easier for others to reuse and cite the research.
* Good documentation enables others—or even the original researchers themselves—to build on the work easily. This ensures that the research continues.

This also means that the whole scientific field benefits!

### Slide 20



Next, we will have a short group activity again.

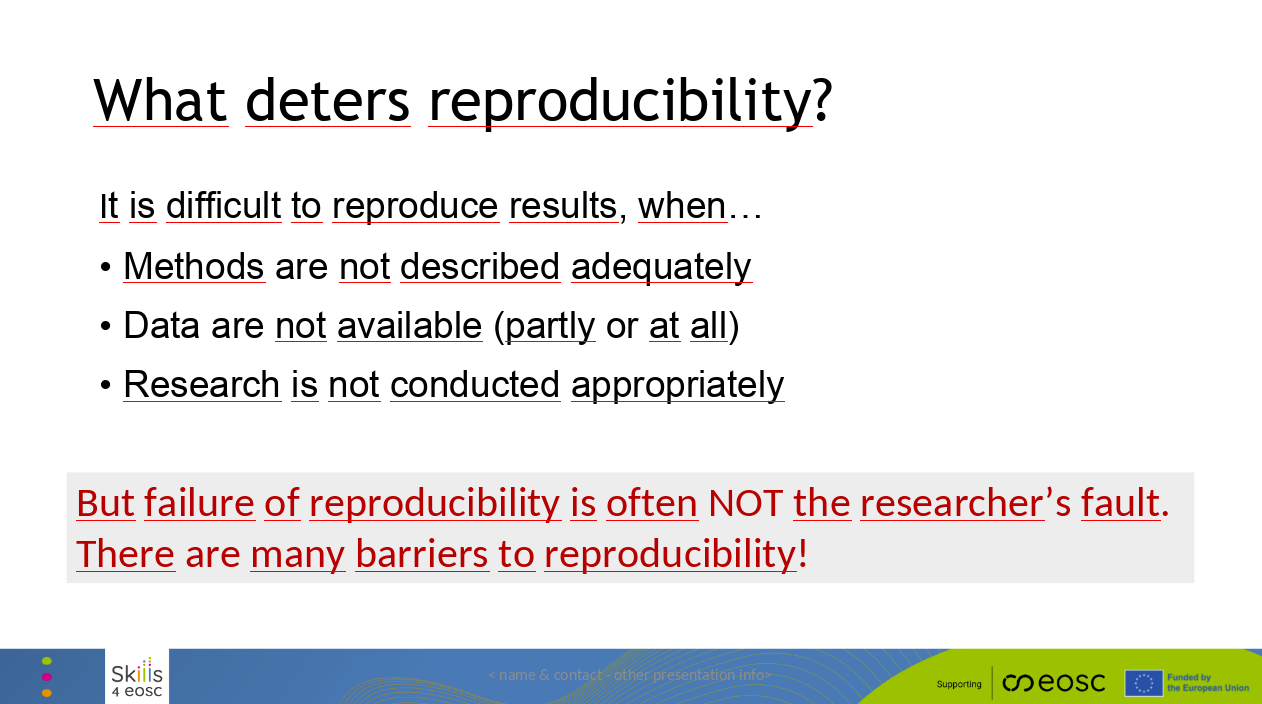
We will divide you into groups and send you into breakout rooms.

Once you are with your group, here are two questions that we would like you to discuss.

In your experience or in your scientific field:

1. Why is reproducibility hard?
2. What are some barriers and challenges?

Slide 21



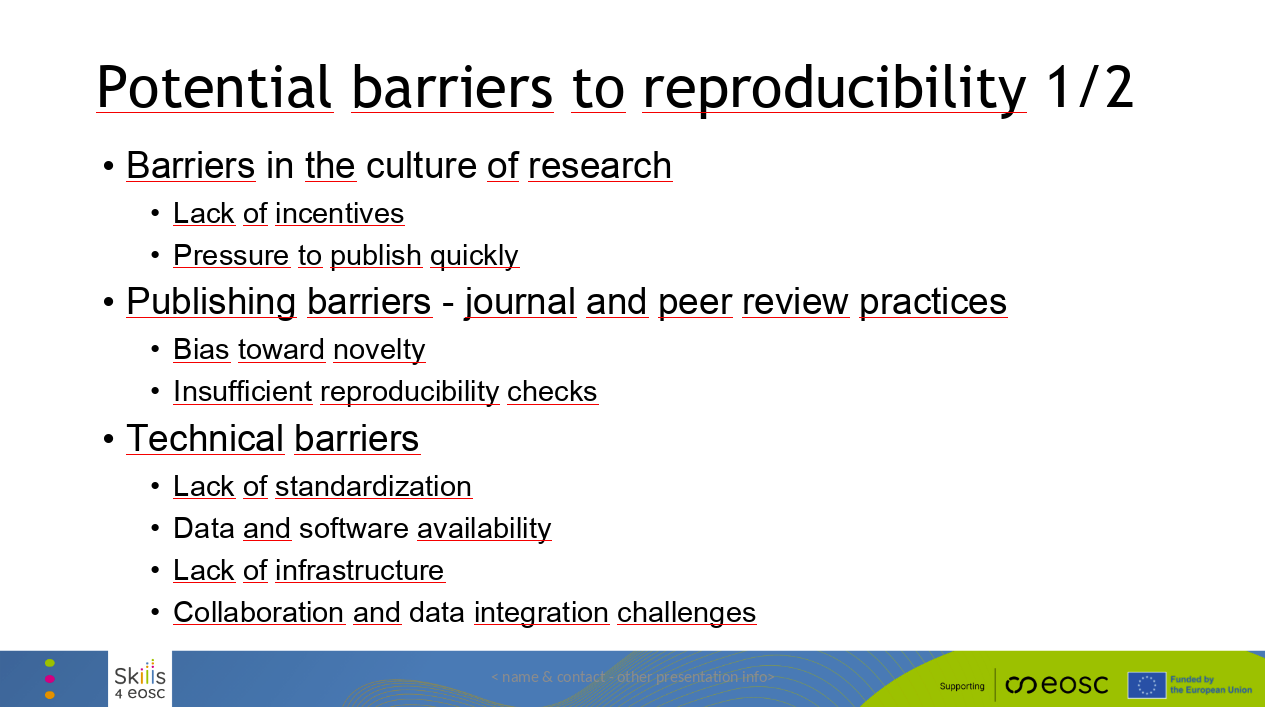
It can be really hard to reproduce results when:

* The methods aren’t described clearly enough,
* The data isn’t available,
* or the research wasn’t done in a way that makes it possible to reproduce it.

These are things that are usually the responsibility of the researcher.

But it’s important to remember that failing to reproduce results isn’t usually the researcher’s fault. There are many barriers that make reproducibility difficult! We will talk about them next.

### Slide 22



Reproducibility has several challenges, which we have here grouped into seven categories. Here are the first three:

**1. Cultural Barriers:**

* Researchers are rewarded for producing high-impact, novel results, not for making their work reproducible.
* The ‘publish or perish’ culture pressures them to prioritize speed over careful validation.
* Success is often measured by quantity and impact, not quality or reproducibility.

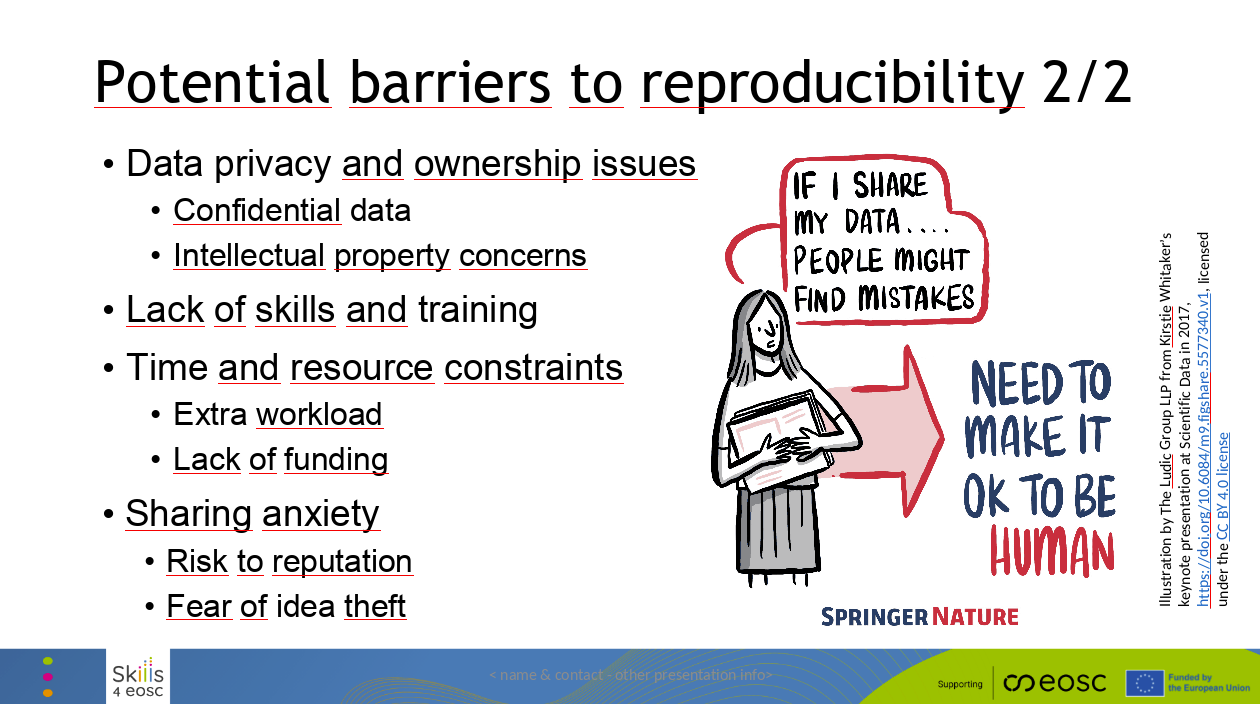
**2.** **Publishing Barriers:**

* Journals favor new findings over replication studies.
* Many journals don’t require authors to share data, code, or methods, and peer reviewers don’t often have the time or expertise to verify reproducibility.

**3. Technical Barriers:**

* (Lack of Standardization) Different fields often use their own methods, tools, and protocols, which makes it harder to reproduce work across studies.
* (Data and Software Availability) Outdated or unavailable software can make it impossible to reproduce results.
* (Lack of Infrastructure) Some fields don’t have reliable platforms to store and share large datasets, and reproducing complex simulations can require resources that not everyone has.
* (Collaboration and Data Integration Challenges) Interdisciplinary work can face unique problems, like differences in data formats, methods, or terminology. Large datasets also come with logistical and technical challenges.

### Slide 23



### And here we have four more barriers:

**Data Privacy and Ownership Issues**:

* Ethical concerns and privacy laws (e.g., patient records) can prevent sharing sensitive data.
* Researchers may avoid sharing their data or methods because they’re worried about losing intellectual property or competitive advantage.

**Lack of Skills and Training:**

* Many researchers lack training in data management, version control, and open science tools.
* Technical tools like version control and data sharing platforms can be complex to use for those who are unfamiliar with them.

**Time and Resource Constraints:**

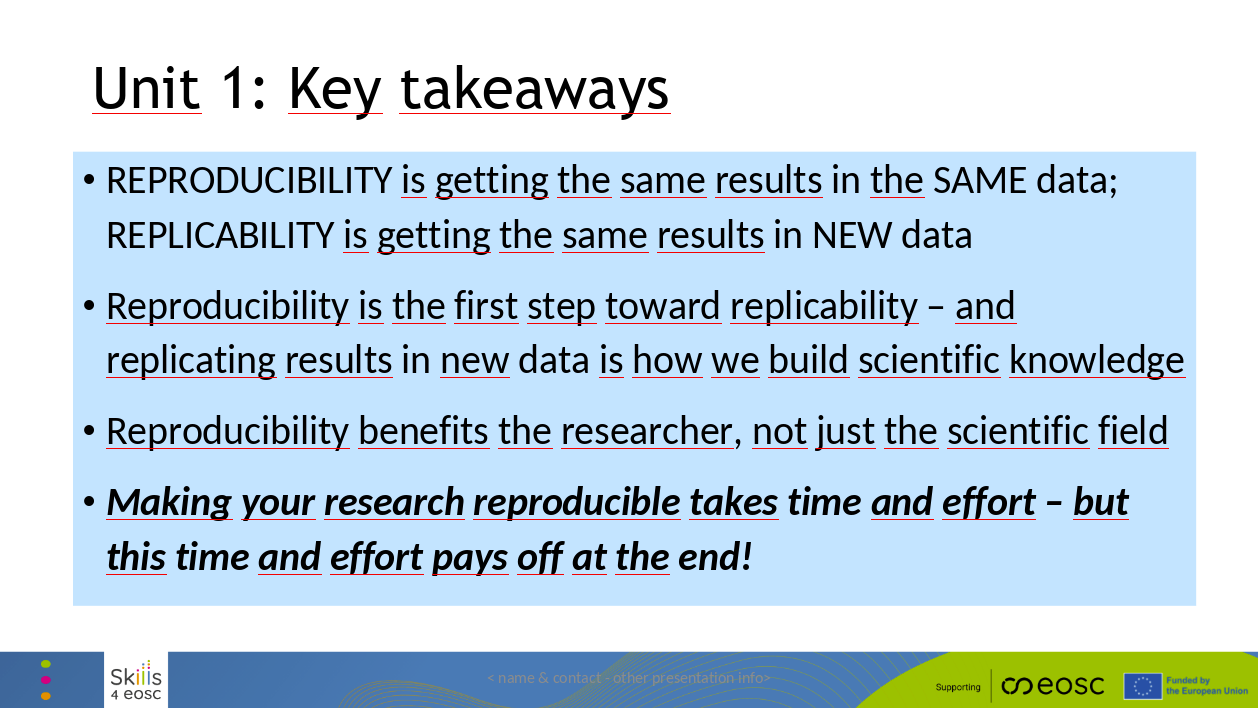
* Ensuring reproducibility requires extra effort in documenting, sharing data, and maintaining records, which can be hard, especially under tight deadlines.
* Research grants often don’t cover the additional costs for reproducibility, such as data storage or validation studies.

**Fears:**

* Researchers may avoid transparency due to fear of exposing errors, which could harm their reputation.
* Researchers may withhold data or methods to protect their ideas from being stolen or misused.

*These barriers show that reproducibility is a systemic issue, not just a responsibility of an individual researcher.*

### Slide 24

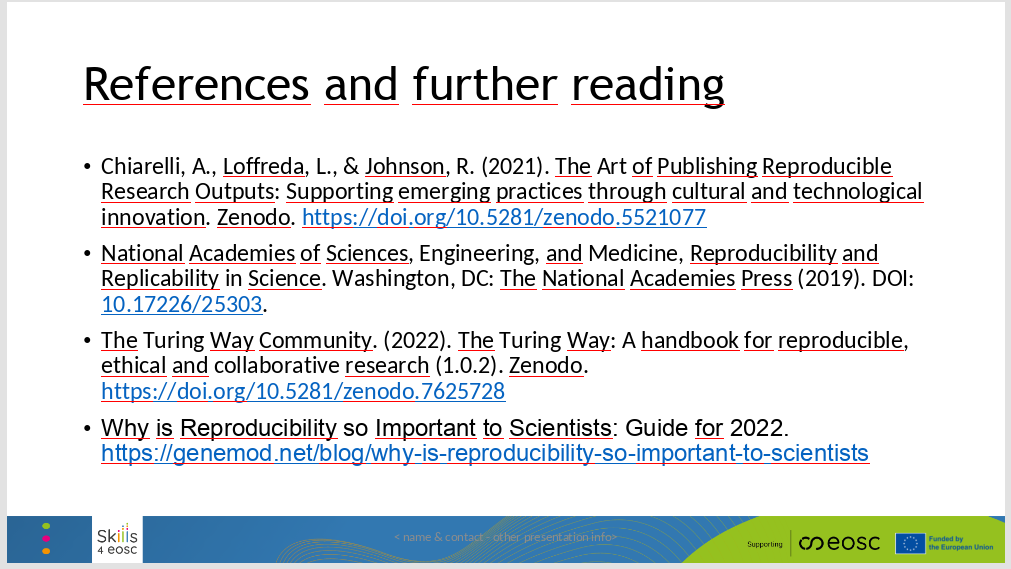


Reproducibility means getting the same results with the same data, while replicability means getting the same results with new data.

Reproducibility is the first step toward replicability, and replicating results is how we build scientific knowledge.

Reproducibility benefits researchers directly, not just the broader scientific community. It takes time and effort, but that investment pays off in the long run!

### Slide 25



Here are some references and further reading on the topic of reproducibility and replicability in scientific research.

### Slide 26



Thank you for your attention and we are happy to take any questions.