

# Integration of Research software into the EOSC infrastructure

Lessons learned from Computer science



UMEÅ UNIVERSITY

WASP | WALLENBERG AI,  
AUTONOMOUS SYSTEMS  
AND SOFTWARE PROGRAM

# /whoami



Mohammad Reza Saleh Sedghpour  
Dpt. Computing science, Umeå University  
msaleh@cs.umu.se  
Twitter: @sedghpour



Sanna Isabel Ulfsparre  
Umeå University Library  
sanna.isabel.ulfsparre@umu.se  
[www.umu.se/personal/sanna-isabel-ulfsparre/](http://www.umu.se/personal/sanna-isabel-ulfsparre/)

# **Research software (RS) in EOSC**

EOSC infrastructure need to:

- include RS to fulfill the FAIR and the FAIR4RS principles
- support sharing, collaborating, evaluating, reproducing, re-using and preserving software
- have a research domain neutral core of standards.

# **Transparency in research and research-related practices**



# **Rigor, credibility and validity**

For qualitative research, it is important to have transparency to:

- show that conclusions and analysis are justifiable
- make it possible to discuss uncounscious biases and ethical issues
- show a clear relationship between data, theory and method, as well as rigour in the execution.

# How do you define "Reproducibility"?

[www.menti.com](https://www.menti.com)

Code: 67 80 83 6





Menti

# Reproducibility

*Reproducibility* is the procedure of **independent confirmation** of a scientific **hypothesis** by **another team**.

(Vitek and Kalibera 2011)

# How do you define "Repeatability"?

[www.menti.com](https://www.menti.com)

Code: 67 80 83 6







Menti

# Repeatability

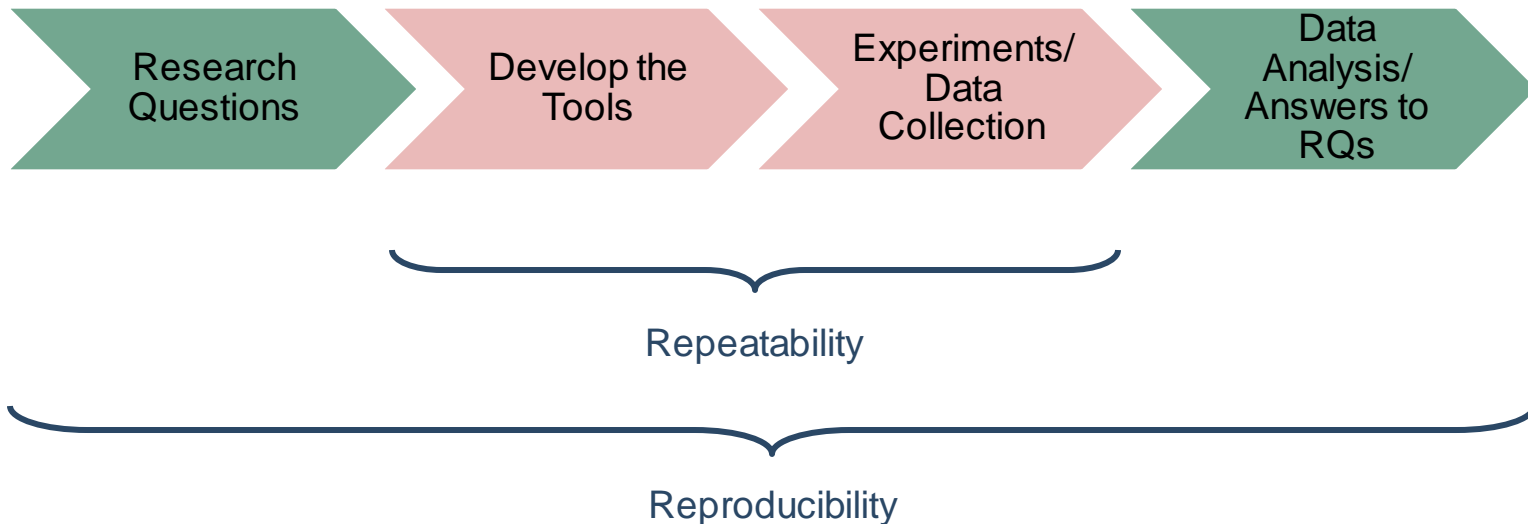
The ability to re-run the exact **same experiment** with the **same procedure** on the **same or comparable system** and receive the **same or very similar result** is referred to as *repeatability*.

(Vitek and Kalibera 2011)



Menti

# Distributed Systems Research



# Is the current research repeatable and reproducible?

[www.menti.com](https://www.menti.com)

Code:67 80 83 6





Menti

# Reproducibility crisis

In a survey published 2016, 90% of researchers answered that there was a **slight** or a **significant** reproducibility crisis.

(Baker, Science 2016)

"[...] the scientific community was shaken by reports that a troubling proportion of peer-reviewed preclinical studies are not reproducible."

(McNutt, Nature 2014)





Menti

## **Conclusion:**

**EOSC need to support RS in order to support reproducibility, repeatability and validation of qualitative processes.**

# Questions during the presentation

[www.menti.com](https://www.menti.com)

Code: 67 80 83 6





Menti

# **EOSC for RS**

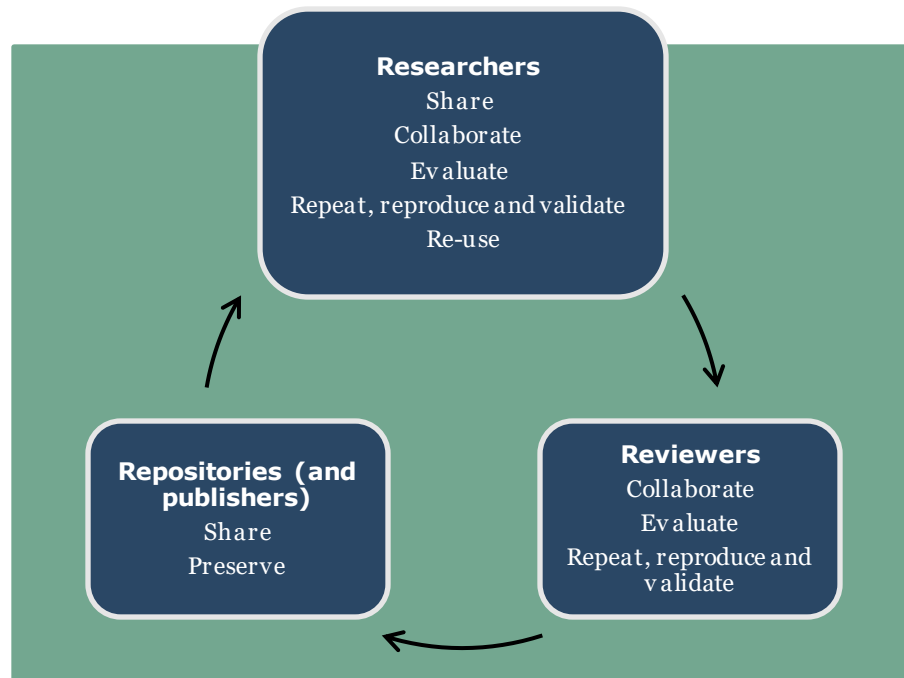
- Share
- Collaborate
- Evaluate
- Reproduce, repeat  
and validate
- Re-use
- Preserve



Menti

# EOSC for RS

- Share
- Collaborate
- Evaluate
- Reproduce, repeat and validate
- Re-use
- Preserve



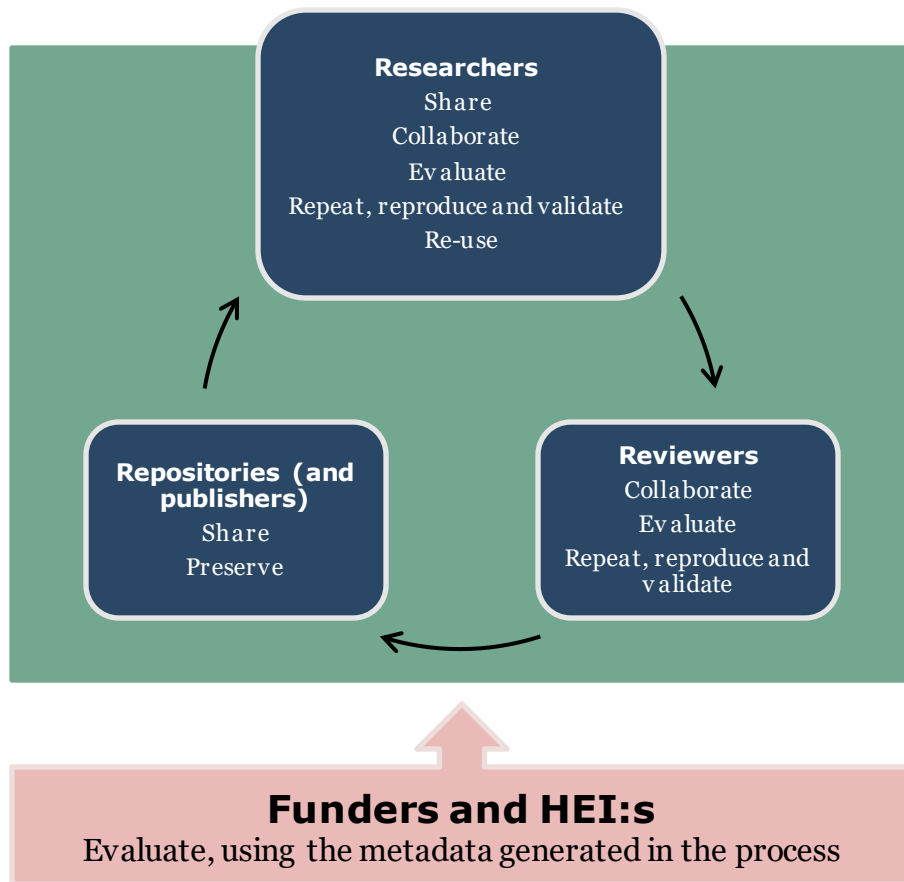




Menti

# EOSC for RS

- Share
- Collaborate
- Evaluate
- Reproduce, repeat and validate
- Re-use
- Preserve





Menti

# Integration of Computer science practices

Generalisation of existing practices for:

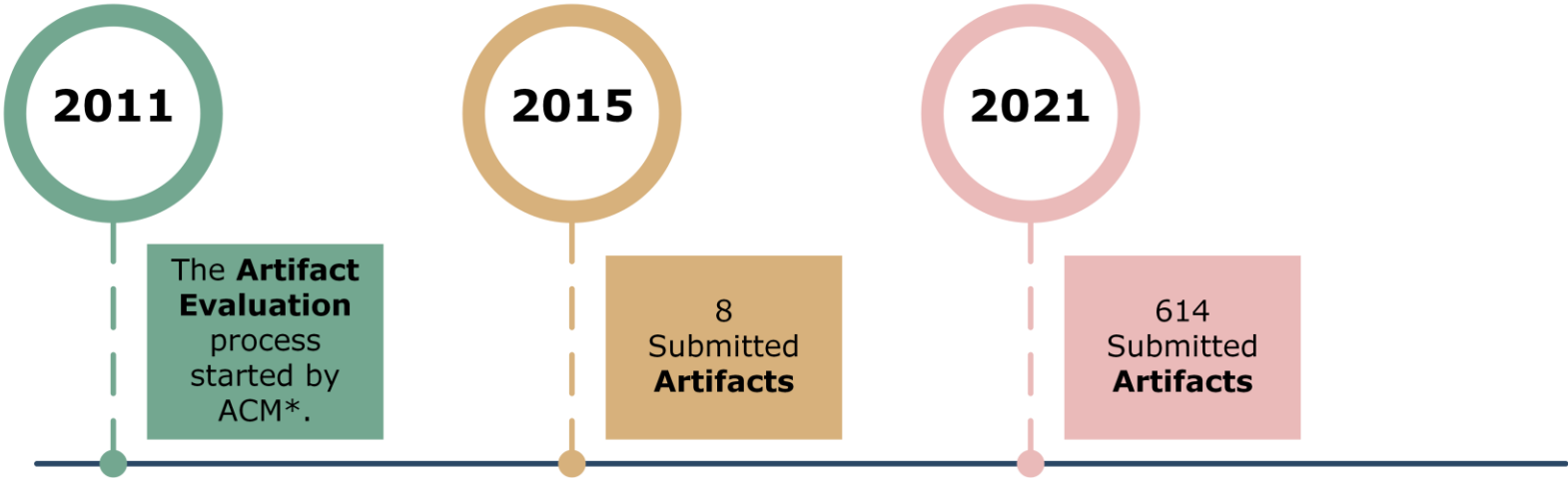
- Open source communities
- Infrastructures and features
- Evaluation and quality control
- Scholarly communication practices





Menti

# The badging system

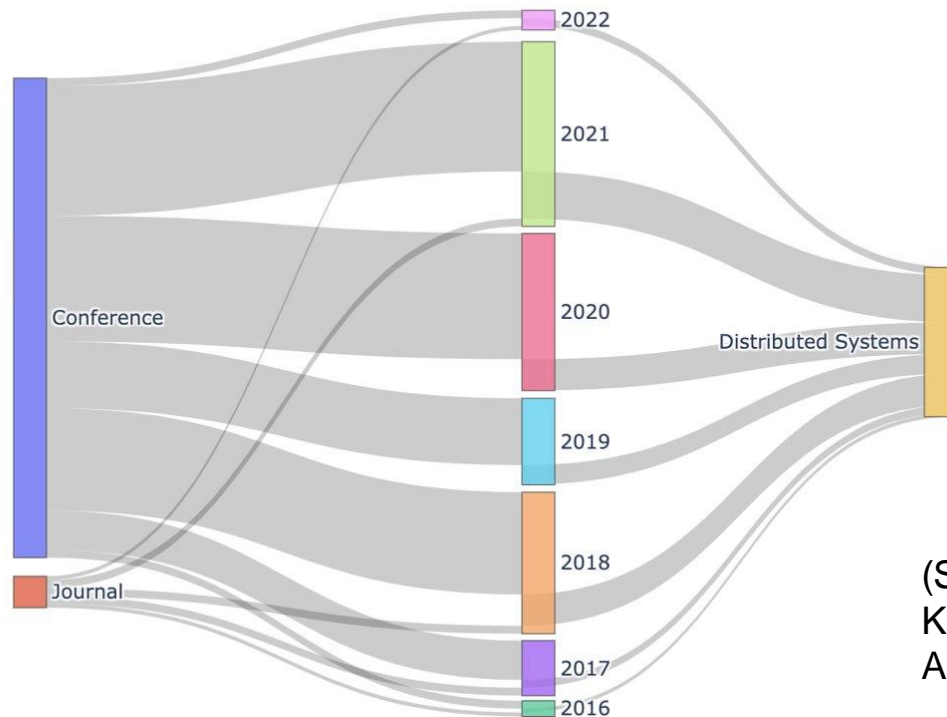


\* Association for Computing Machinery



Menti

# Timeline for artifact evaluation in distributed systems



(Saleh Sedghpour, M.R;  
Klein, C.; Papadopoulos,  
A.V. *et al.* Est. 2023.)



Menti

# The study

- ACM Digital Library
- IEEE Xplore
- WikiCFP
- Researchr

No	Keyword
1	Distributed Systems
2	Cloud Computing
3	Edge Computing
4	Fog Computing
5	Serverless Computing
6	Service-Oriented Architectures
7	Microservices



Menti

# The study: Process

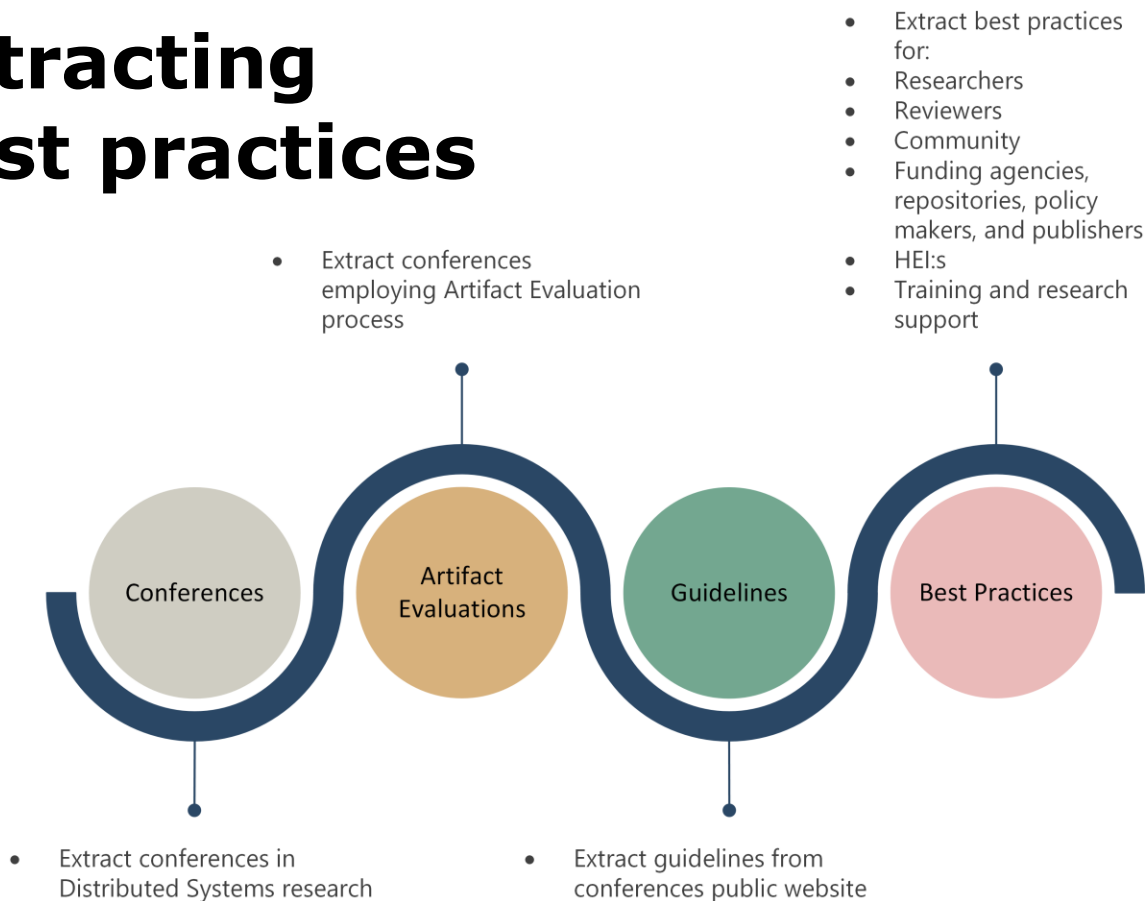
Conference Name	Years
ASPLOS	2022, 2021, 2020
CF	2020
CoNEXT	2022, 2021, 2020
ESEC/FSE	2021, 2020
EuroSys	2022, 2021
ICPE	2022, 2021, 2020
ICSE	2022, 2021, 2020

Conference Name	Years
MICRO	2021
Middleware	2022, 2021, 2020
SC	2022, 2021, 2020
SOSP	2021
PPoPP	2022, 2021, 2020
UCC	2021



Menti

# Extracting best practices





# Best practices: Researchers

Introduction





# Best practices: Researchers

Introduction	Archival Repository (DOI)
--------------	---------------------------------



# Best practices: Researchers

Introduction	Archival Repository (DOI)	Algorithms
--------------	---------------------------------	------------



Menti

# Best practices: Researchers

Introduction	Archival Repository (DOI)	Algorithms	Required Resources
--------------	---------------------------------	------------	-----------------------



Menti

# Best practices: Researchers

Introduction	Archival Repository (DOI)	Algorithms	Required Resources
Inputs			



Menti

# Best practices: Researchers

Introduction	Archival Repository (DOI)	Algorithms	Required Resources
Inputs	Run-Time States		



Menti

# Best practices: Researchers

Introduction	Archival Repository (DOI)	Algorithms	Required Resources
Inputs	Run-Time States	Problem Dimensions	



Menti

# Best practices: Researchers

Introduction	Archival Repository (DOI)	Algorithms	Required Resources
Inputs	Run-Time States	Problem Dimensions	Metrics



Menti

# Best practices: Researchers

Introduction	Archival Repository (DOI)	Algorithms	Required Resources
Inputs	Run-Time States	Problem Dimensions	Metrics
Planning			





# Best practices: Researchers

Introduction	Archival Repository (DOI)	Algorithms	Required Resources
Inputs	Run-Time States	Problem Dimensions	Metrics
Planning	Expected Outputs		



Menti

# Best practices: Researchers

Introduction	Archival Repository (DOI)	Algorithms	Required Resources
Inputs	Run-Time States	Problem Dimensions	Metrics
Planning	Expected Outputs	Sample Data	



Menti

# Best practices: Researchers

Introduction	Archival Repository (DOI)	Algorithms	Required Resources
Inputs	Run-Time States	Problem Dimensions	Metrics
Planning	Expected Outputs	Sample Data	Licenses



Menti

# Best practices: Researchers

Introduction	Archival Repository (DOI)	Algorithms	Required Resources	Programs
Inputs	Run-Time States	Problem Dimensions	Metrics	
Planning	Expected Outputs	Sample Data	Licenses	



Menti

# Best practices: Researchers

Introduction	Archival Repository (DOI)	Algorithms	Required Resources	Programs	Experiment
Inputs	Run-Time States	Problem Dimensions	Metrics		
Planning	Expected Outputs	Sample Data	Licenses		



Menti

# Best practices: Researchers

Introduction	Archival Repository (DOI)	Algorithms	Required Resources	Programs	Experiment
Inputs	Run-Time States	Problem Dimensions	Metrics	Source Control	
Planning	Expected Outputs	Sample Data	Licenses		



Menti

# Best practices: Researchers

Introduction	Archival Repository (DOI)	Algorithms	Required Resources	Programs	Experiment
Inputs	Run-Time States	Problem Dimensions	Metrics	Source Control	Automation
Planning	Expected Outputs	Sample Data	Licenses		



Menti

# Best practices: Researchers

Introduction	Archival Repository (DOI)	Algorithms	Required Resources	Programs	Experiment
Inputs	Run-Time States	Problem Dimensions	Metrics	Source Control	Automation
Planning	Expected Outputs	Sample Data	Licenses	Test Driven Developing	





Menti

# Best practices: Researchers

Introduction	Archival Repository (DOI)	Algorithms	Required Resources	Programs	Experiment
Inputs	Run-Time States	Problem Dimensions	Planning	Source Control	Automation
Metrics	Expected Outputs	Sample Data	Licenses	Test Driven Developing	Public Cloud



Menti

# Best practices: Researchers

Metadata describing software				Software	
Introduction	Archival Repository (DOI)	Algorithms	Required Resources	Programs	Experiment
Inputs	Run-Time States	Problem Dimensions	Metrics	Source Control	Automation
Planning	Expected Outputs	Sample Data	Licenses	Test Driven Developing	Public Cloud



Menti

# Best practices: Reviewers



Multi-stage reviewing



Menti

# Best practices: Reviewers



Multi-stage reviewing



Providing guidelines and checklists



Menti

# Best practices: Reviewers



Multi-stage reviewing



Providing guidelines and checklists



Communications between reviewers should be facilitated and encouraged



Menti

# Best practices: Reviewers



Multi-stage reviewing



Providing guidelines and checklists



Communications between reviewers should be facilitated and encouraged



Access to a public cloud, such as EOSC



Menti

# Best practices: Research communities



Research and discussions on best practices



Menti

# Best practices: Research communities



Research and discussions on best practices



Encouragement of and engagement in research software reviews





Menti

# Best practices: Research communities



Research and discussions on best practices



Encouragement of and engagement in research software reviews



Evaluation of new ideas and methods, such as crowd-sourcing and citizen science



Menti

# Best practices: Research communities



Research and discussions on best practices



Encouragement of and engagement in research software reviews



Evaluation of new ideas and methods, such as crowd-sourcing and citizen science



Inter-disciplinary working groups for research software practices



Menti

# Best practices: Research communities



Research and discussions on best practices



Encouragement of and engagement in research software reviews



Evaluation of new ideas and methods, such as crowd-sourcing and citizen science



Inter-disciplinary working groups for research software practices



Support and exploration of practices that follow the FAIR principles

# **Our recommendation: Policy makers, publishers and funding agencies**

- ✓ FAIR research software practices should be embedded in policy and funding requirements.



# **Our recommendation:** **Policy makers, publishers** **and funding agencies**

- ✓ FAIR research software practices should be embedded in policy and funding requirements.
- ✓ Generous metadata describing software should be expected.





# Our recommendation: Policy makers, publishers and funding agencies

- ✓ FAIR research software practices should be embedded in policy and funding requirements.
- ✓ Generous metadata describing software should be expected.
- ✓ Access to research software should be encouraged "as openly as possible, as restricted as necessary".



# **Our recommendation:** **Policy makers, publishers** **and funding agencies**

- ✓ FAIR research software practices should be embedded in policy and funding requirements.
- ✓ Generous metadata describing software should be expected.
- ✓ Access to research software should be encouraged "as openly as possible, as restricted as necessary".
- ✓ Research on open science and practices for reproducibility should be supported and funded.





# Our recommendation:

## Policy makers, publishers and funding agencies

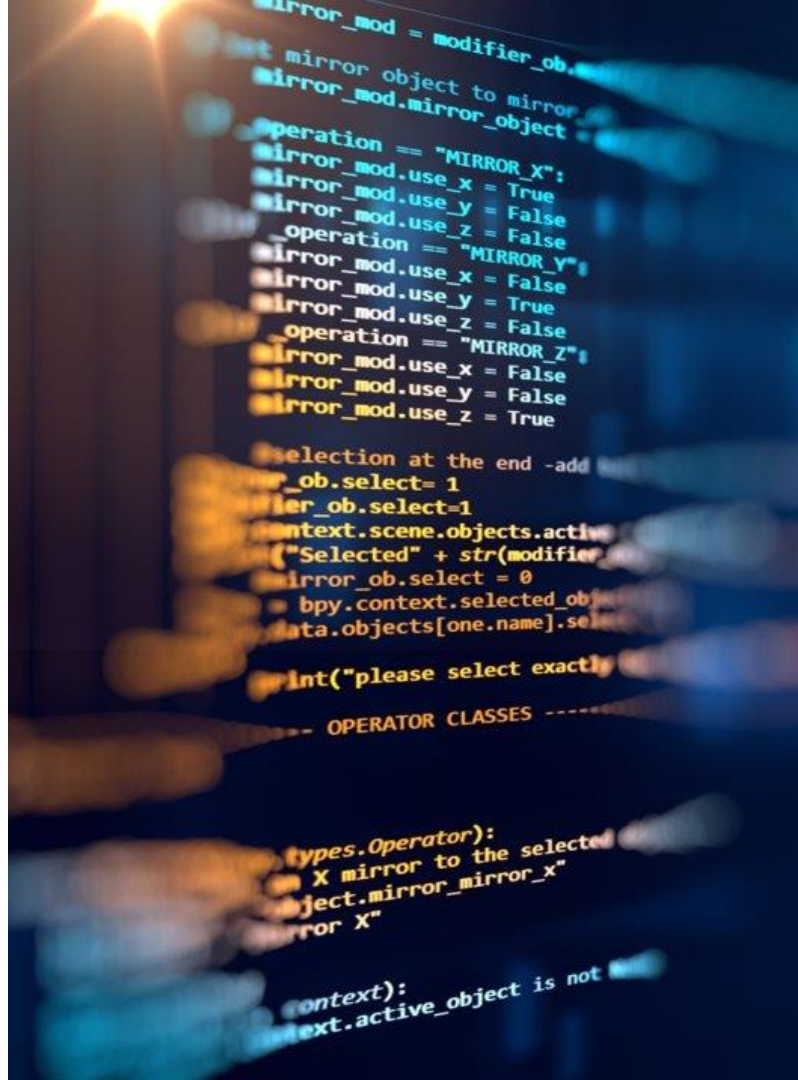
- ✓ FAIR research software practices should be embedded in policy and funding requirements.
- ✓ Generous metadata describing software should be expected.
- ✓ Access to research software should be encouraged "as openly as possible, as restricted as necessary".
- ✓ Research on open science and practices for reproducibility should be supported and funded.
- ✓ There should be badging system for software quality that is controlled centrally by an entity such as EOSC.





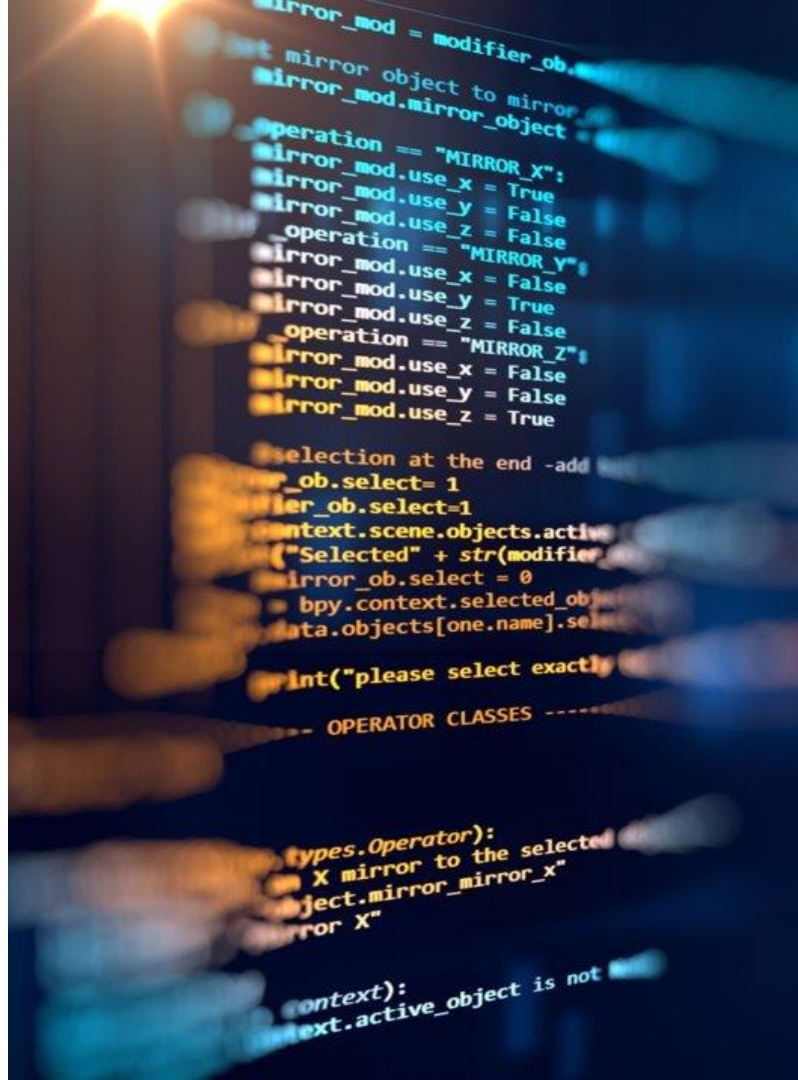
# Our recommendation: Training and researcher support

- Basic programming and "Clean code"



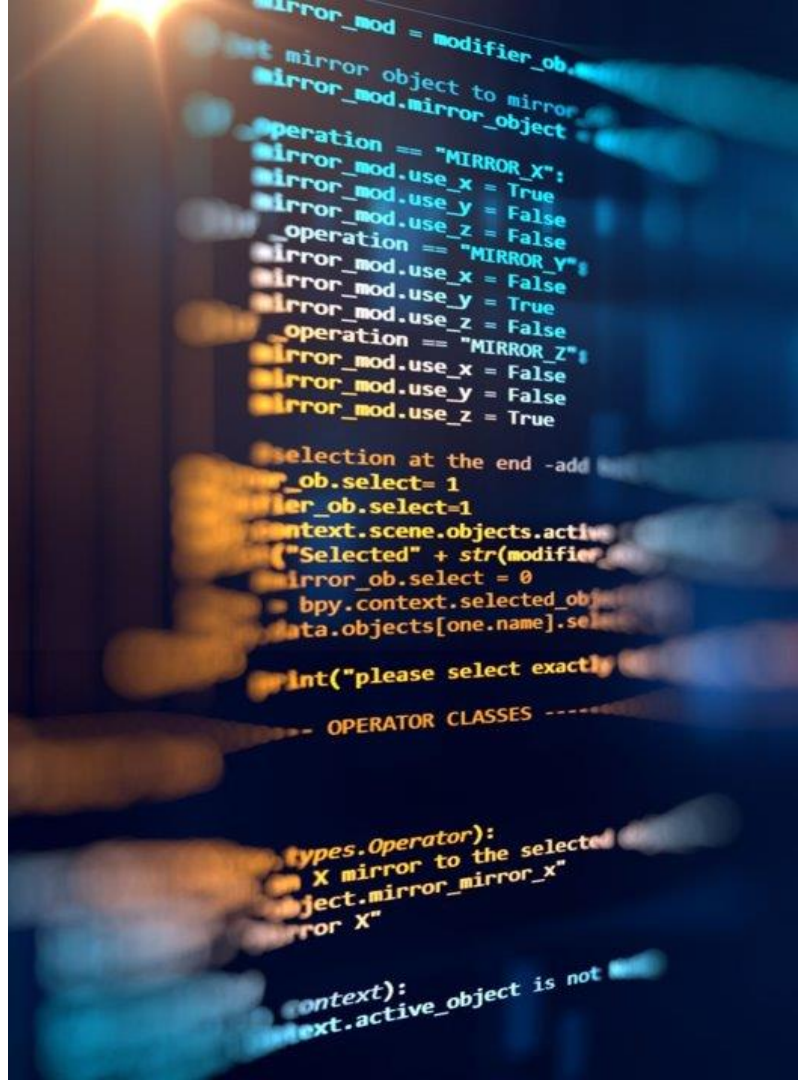
# Our recommendation: Training and researcher support

- Basic programming and "Clean code"
- Development methods:
  - Design patterns
  - Automation
  - Test driven development



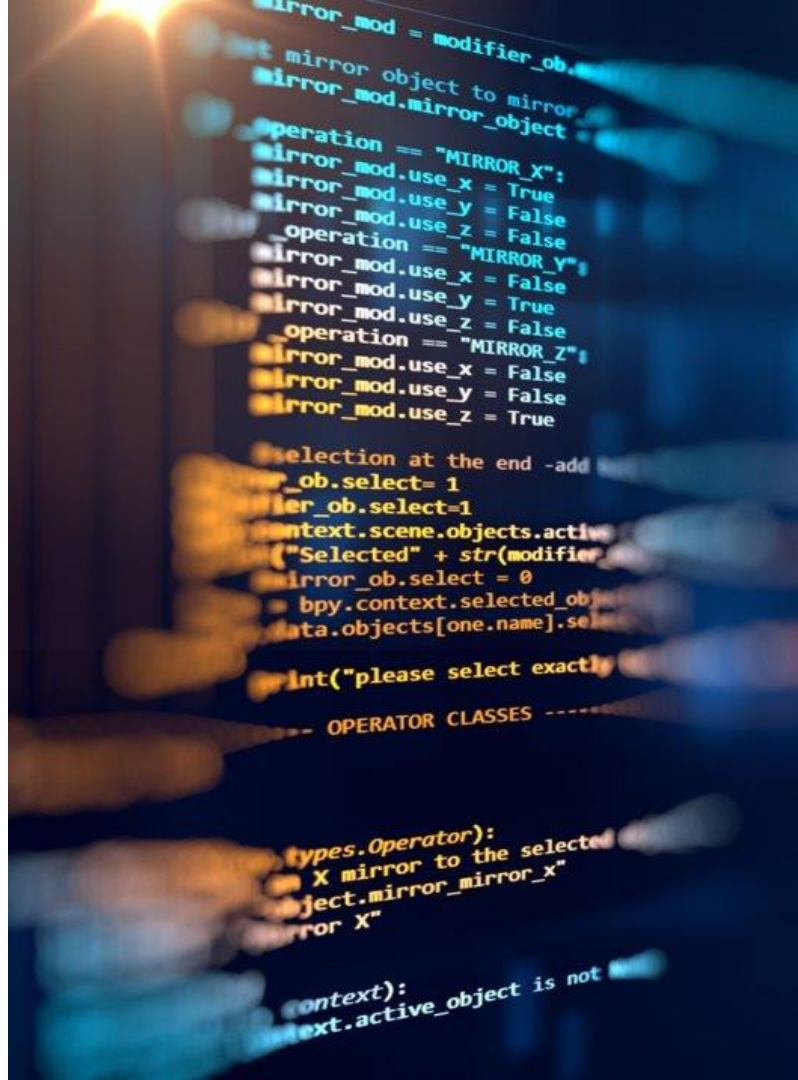
# Our recommendation: Training and researcher support

- Basic programming and "Clean code"
- Development methods:
  - Design patterns
  - Automation
  - Test driven development
- Iterative thinking



# Our recommendation: Training and researcher support

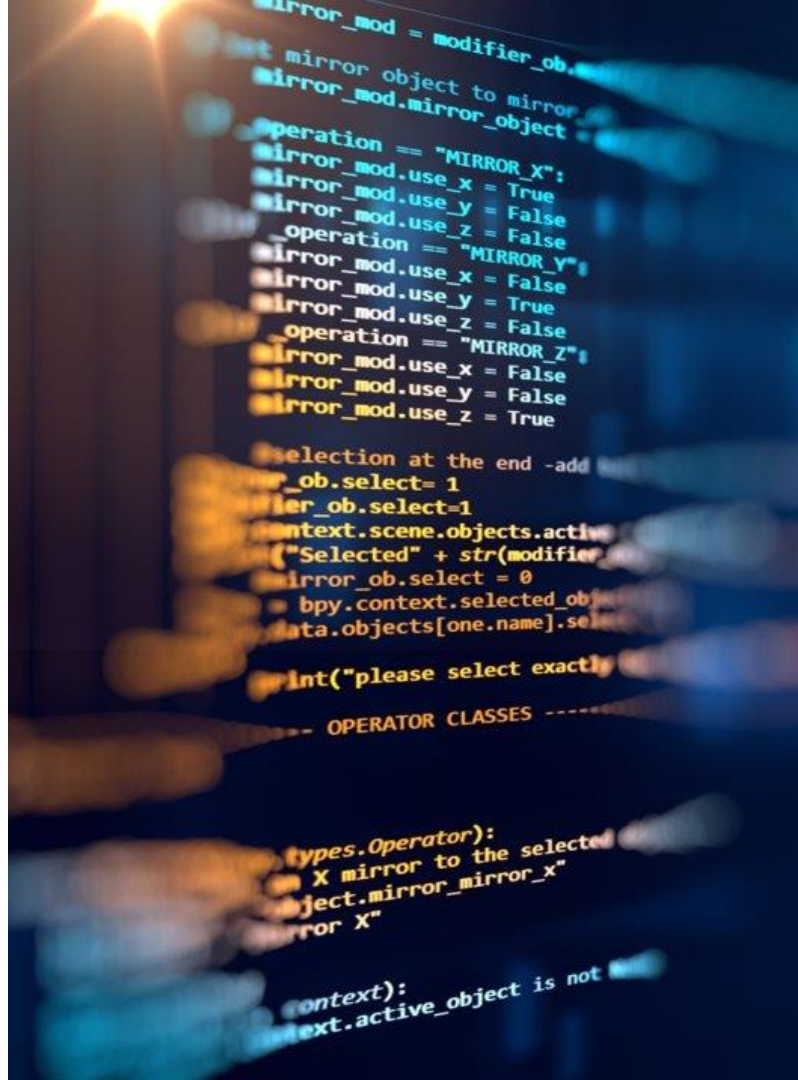
- Basic programming and "Clean code"
- Development methods:
  - Design patterns
  - Automation
  - Test driven development
- Iterative thinking
- Use of infrastructures





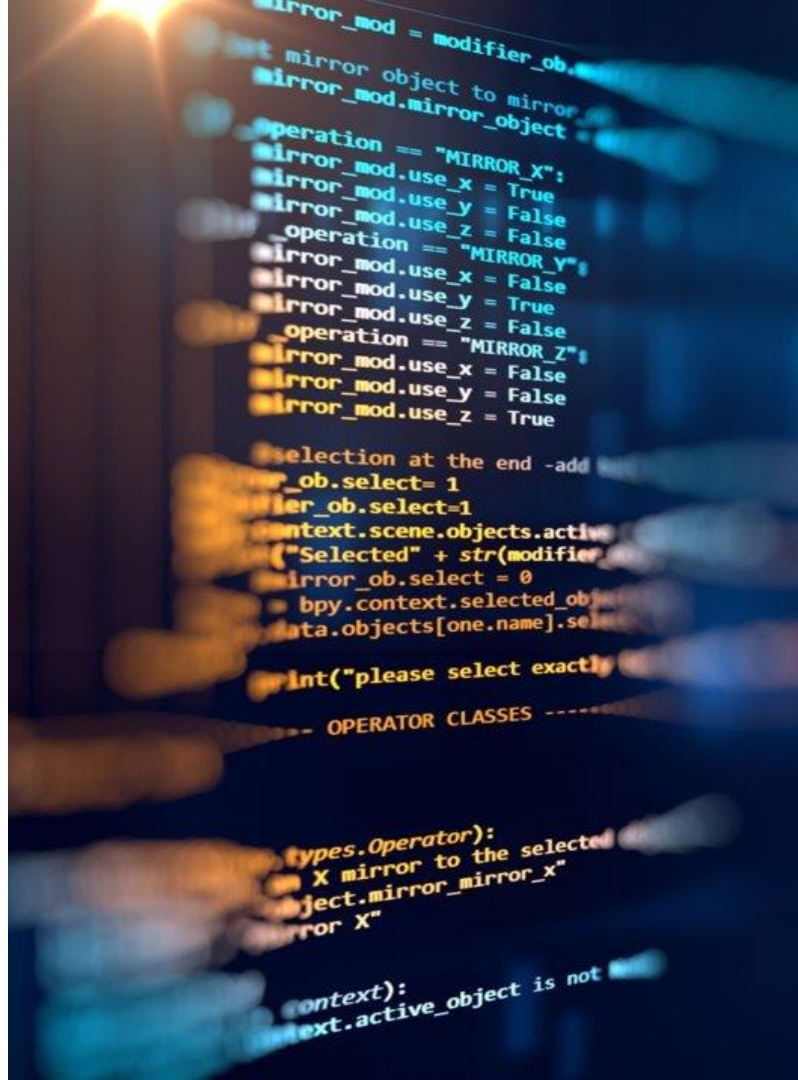
# Our recommendation: Training and researcher support

- Basic programming and "Clean code"
- Development methods:
  - Design patterns
  - Automation
  - Test driven development
- Iterative thinking
- Use of infrastructures
- Research software in Open science and Scholarly communication



# Our recommendation: Training and researcher support

- Basic programming and "Clean code"
- Development methods:
  - Design patterns
  - Automation
  - Test driven development
- Iterative thinking
- Use of infrastructures
- Research software in Open science and Scholarly communication
- Legal training and support



# References

Wilkinson, M.; Dumontier, M.; Aalbersberg, I. *et al.* 2016. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* 3(160018). DOI: [10.1038/sdata.2016.18](https://doi.org/10.1038/sdata.2016.18)

Barker, M.; Chue Hong, N.P.; Katz, D.S. *et al.* 2022. Introducing the FAIR Principles for research software. *Scientific Data* 9(622). DOI: [10.1038/s41597-022-01710-x](https://doi.org/10.1038/s41597-022-01710-x)

Vitek, J. and Kalibera T. 2011. Repeatability, reproducibility and rigor in systems research. *9th ACM International Conference on Embedded Software (EMSOFT)*. Taipei, Taiwan October 9-14 DOI: [10.1145/2038642.2038650](https://doi.org/10.1145/2038642.2038650)

McNutt, M. 2014. Reproducibility. *Science*. 343(6168). DOI: [10.1126/science.1250475](https://doi.org/10.1126/science.1250475)

Baker, M. 2016. 1,500 scientists lift the lid on reproducibility. *Nature* 533: 452–454. DOI: [10.1038/533452a](https://doi.org/10.1038/533452a)

Saleh Sedghpour, M.R; Klein, C.; Papadopoulos, A.V. *et al.* Estimated publishing 2023. Title TBD.



# Thank you!

Mohammad Reza Saleh Sedghpour  
Sanna Isabel Ulfsparre



UMEÅ UNIVERSITY

WASP | WALLENBERG AI,  
AUTONOMOUS SYSTEMS  
AND SOFTWARE PROGRAM