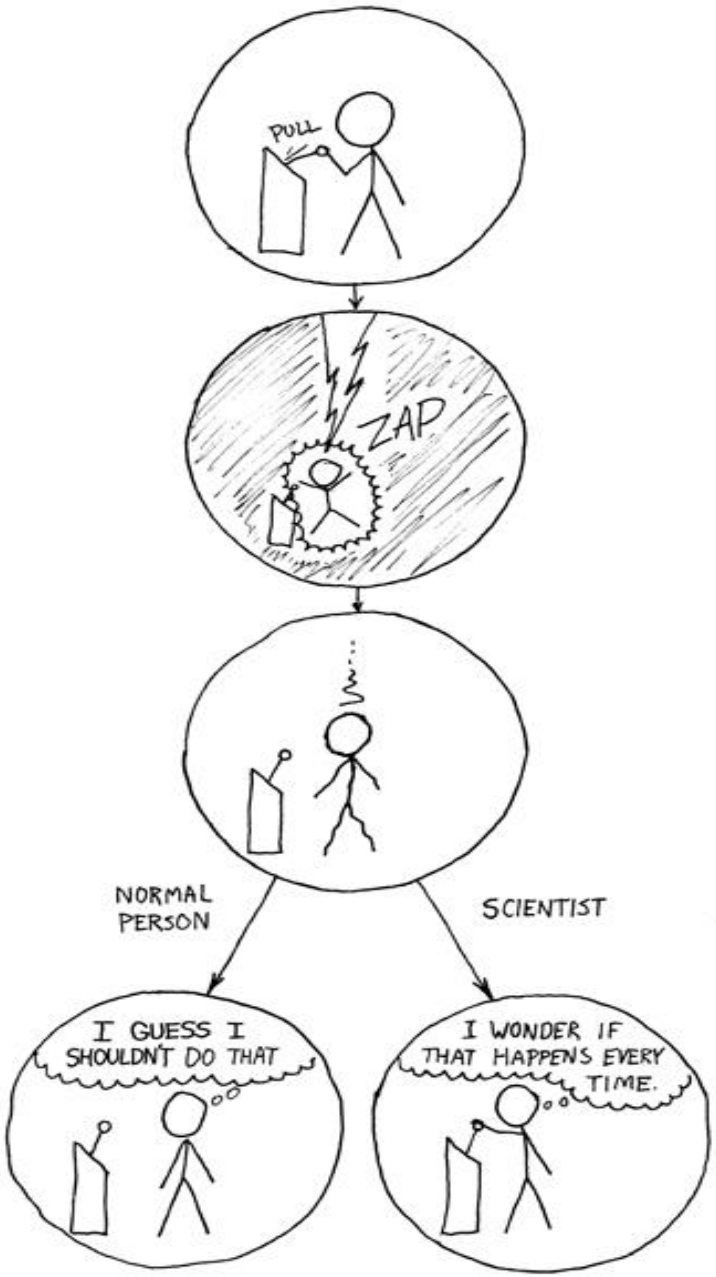


External Certification of Research Reproducibility

Christophe Pérignon, HEC Paris

cascad

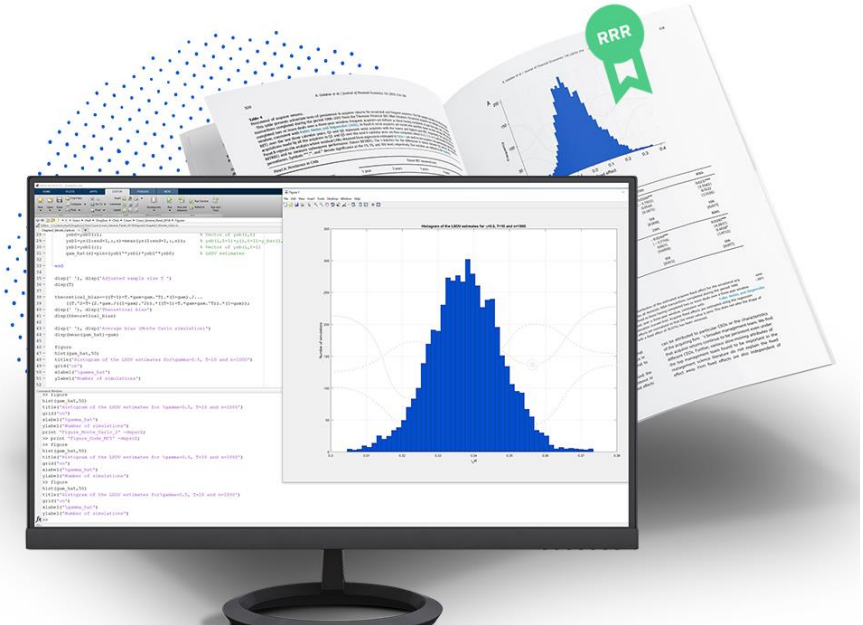
February 13, 2020

ONS, London

Reproducibility certification

cascad = Certification Agency for Scientific Code and Data

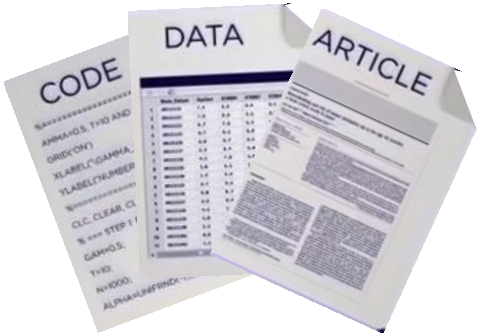
The cascad reproducibility certification attests that the numerical results reported in a scientific article can be reproduced from a set of numerical resources (code and data) provided by the authors.



- www.cascad.tech
- Non-profit academic initiative
- Founded by researchers
- CNRS, HEC Paris, U. Orléans

cascad certification process

Step 1: Submission of the paper and digital resources (code + data)



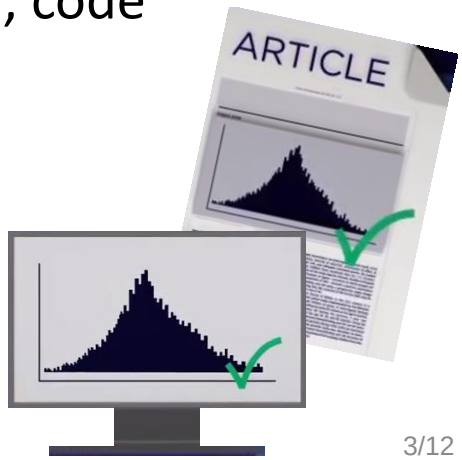
Step 2: Conformity check:

- **cascad** staff verifies whether the submitted resources comply with the **guidelines**
- These guidelines cover the presentation/structure of code and data and aim to make these resources interpretable and reusable
- They are similar to those recommended by the AEA Data Editor and the Social Science Data Editors
- No certification in case of non-conformity

cascad certification process (2)

Step 3: Reviewing process

- **cascad** assigns a **reproducibility reviewer**
- Runs the code and compares the outputs from the code to the numerical results of the paper
- Produces an **execution report**: data description, code description, replication steps, and findings
- Suggests a reproducibility rating



Reproducibility rating

Grade	Comments
RRR	Perfect • All tables and figures, including those in the appendix can be replicated precisely.
RR	Practically Perfect • Whenever complete instructions are available, replicated and published results are equivalent. • Some figures or tables may be time consuming or difficult to reproduce due to missing instructions. However, these exhibits should be approximated and though results may differ slightly, there is no reason to doubt their accuracy.
R	Minor Discrepancies • Minor discrepancies exist between the replication and published replicated results that would not affect the conclusions of the paper (if they were truly errors). • e.g. Estimated coefficients match but standard errors are slightly different and do not substantially change the results of important hypothesis tests. • Difference between RR and R could be the result of computer program versions and operating systems.
D	Potentially Serious Discrepancies • Differences exist between published and replicated results that may indicate problems with the empirical analysis.
DD	Serious Discrepancies • Differences exist between published and replicated results that may indicate problems with the empirical analysis. Substantial differences exist between important empirical results that cannot be reconciled. • These discrepancies indicate that an error in the analysis has probably lead to incorrect conclusions.

Source: Glandon (AER, 2011)

cascad

Execution Report

Title: **Technology Boom, Labor Reallocation, and Human Capital Depreciation**

Authors: **Johan Hombert and Adrien Matray**

Full reference: *Hombert, Johan and Adrien Matray (2019) "Technology Boom, Labor Reallocation, and Human Capital Depreciation", Working Paper.*

The structure and contents of this execution report provided by **cascad** for the certification are similar to those recommended by the AEA Data Editor.

1. DATA DESCRIPTION

This study uses administrative data on French workers and firms, collected by the national statistical office based on a mandatory employer report of the gross earnings of each employee subject to payroll taxes. The dataset includes all employed individuals in the private sector, with information about the gross and net wage, dated employment periods, number of hours worked, job occupation, and the individual's birth year and sex. The data also includes unique firm and establishment identifiers that

Execution report

TABLE 4: QUANTILES OF WAGE GROWTH

Original :

	Wage growth quantiles				
	(1)	(2)	(3)	(4)	(5)
ICT_0	-.105***	-.105***	-.107***	-.121***	-.110***
	(.027)	(.018)	(.015)	(.018)	(.028)
Observations	4,972	4,972	4,972	4,972	4,972

Reproduced :

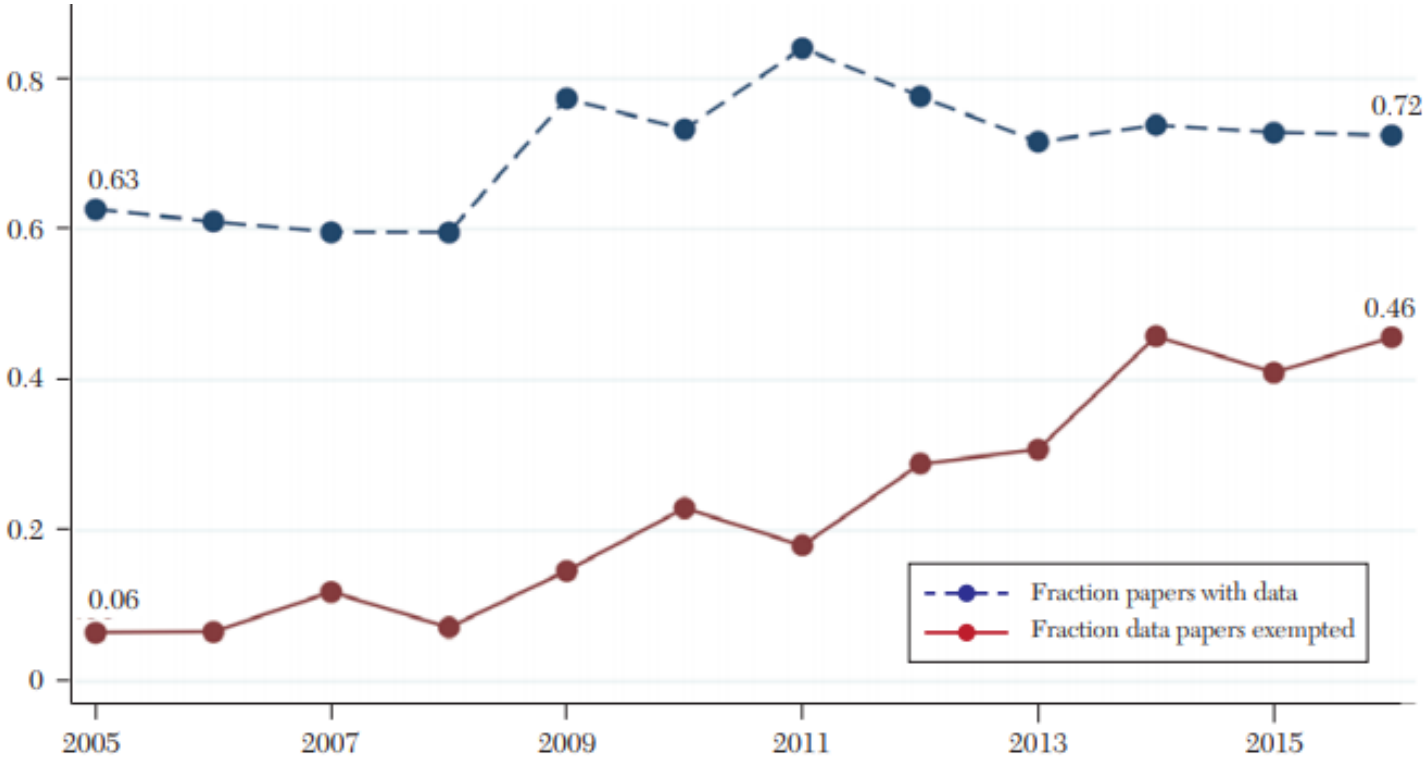
	Wage growth quantiles				
	est1	est2	est3	est4	est5
ICT_0	-0.105***	-0.105***	-0.107***	-0.121***	-0.110***
	(0.025)	(0.017)	(0.015)	(0.016)	(0.031)
Observations	4,972	4,972	4,972	4,972	4,972

cascad certification process (3)

Step 4: Certification

- The **reproducibility editor** assigns the final rating given the results reported in the execution report
- The **reproducibility certificate** and the execution report are sent to the corresponding author
- The cascad staff uploads the certified code and data on an **open access** repository (Zenodo, ICPSR) => **DOI**

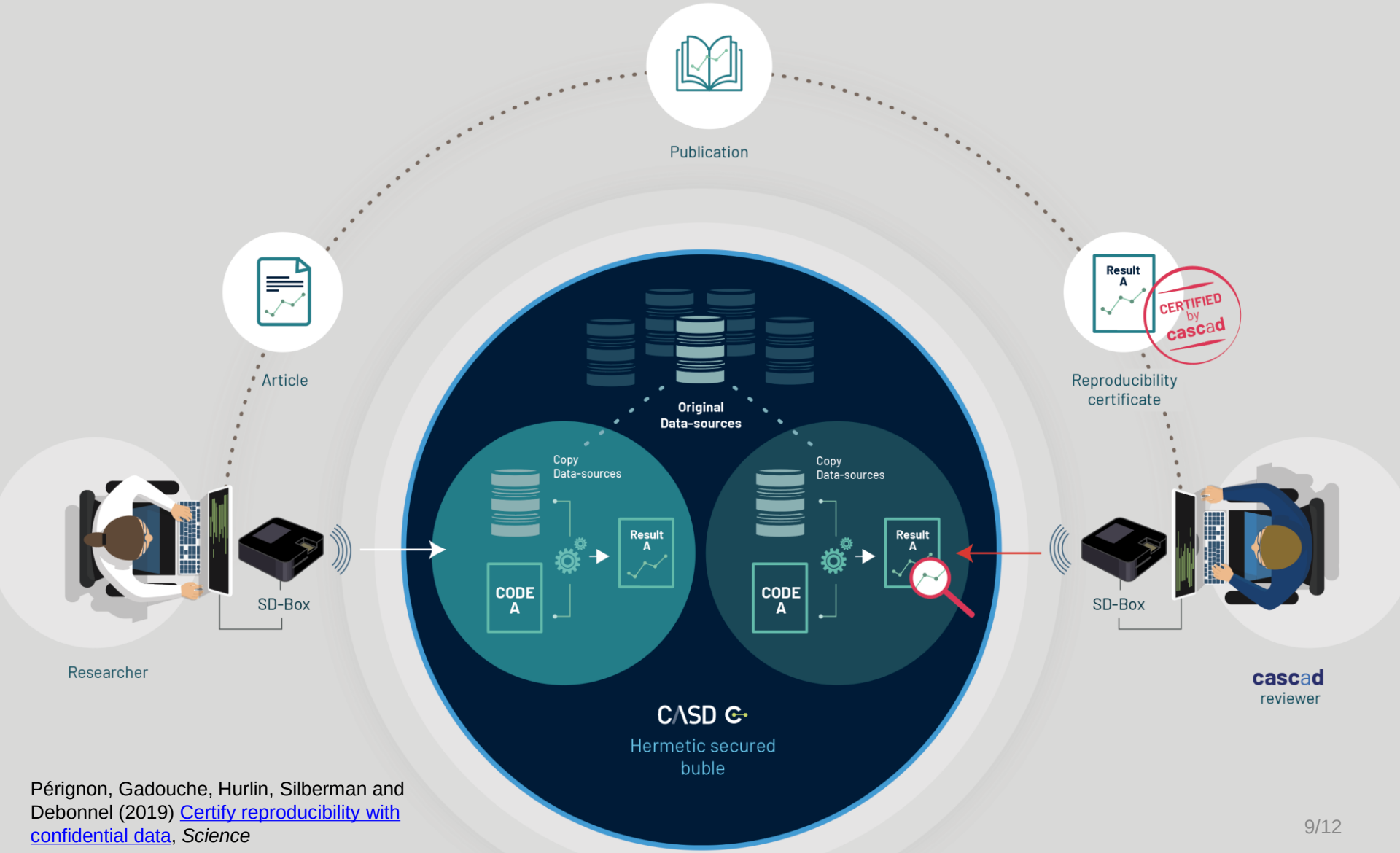
Confidential data



Source: Christensen and Miguel (2018)

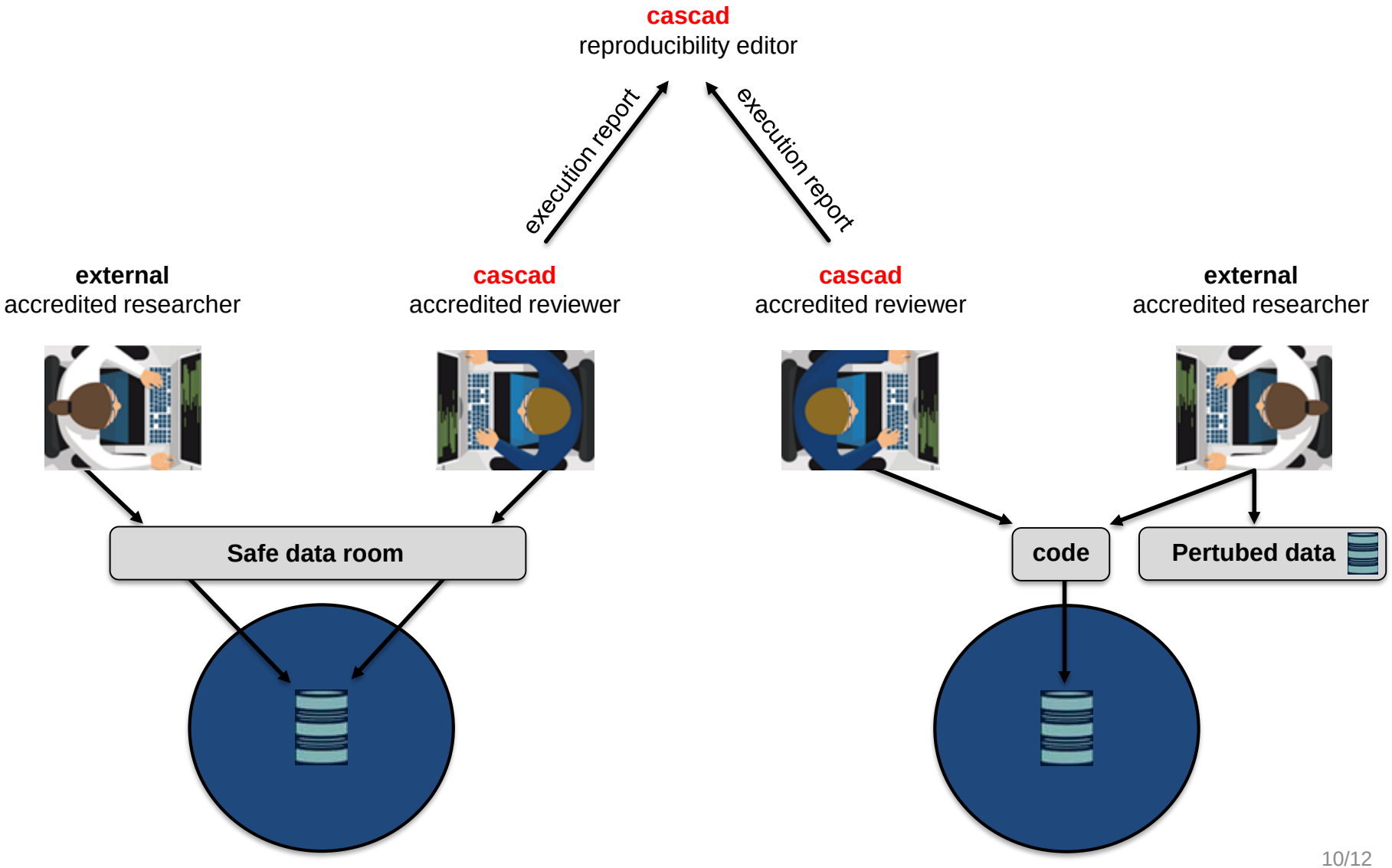
Numerical resources cannot often be shared for confidentiality reasons

Partnership with CASD

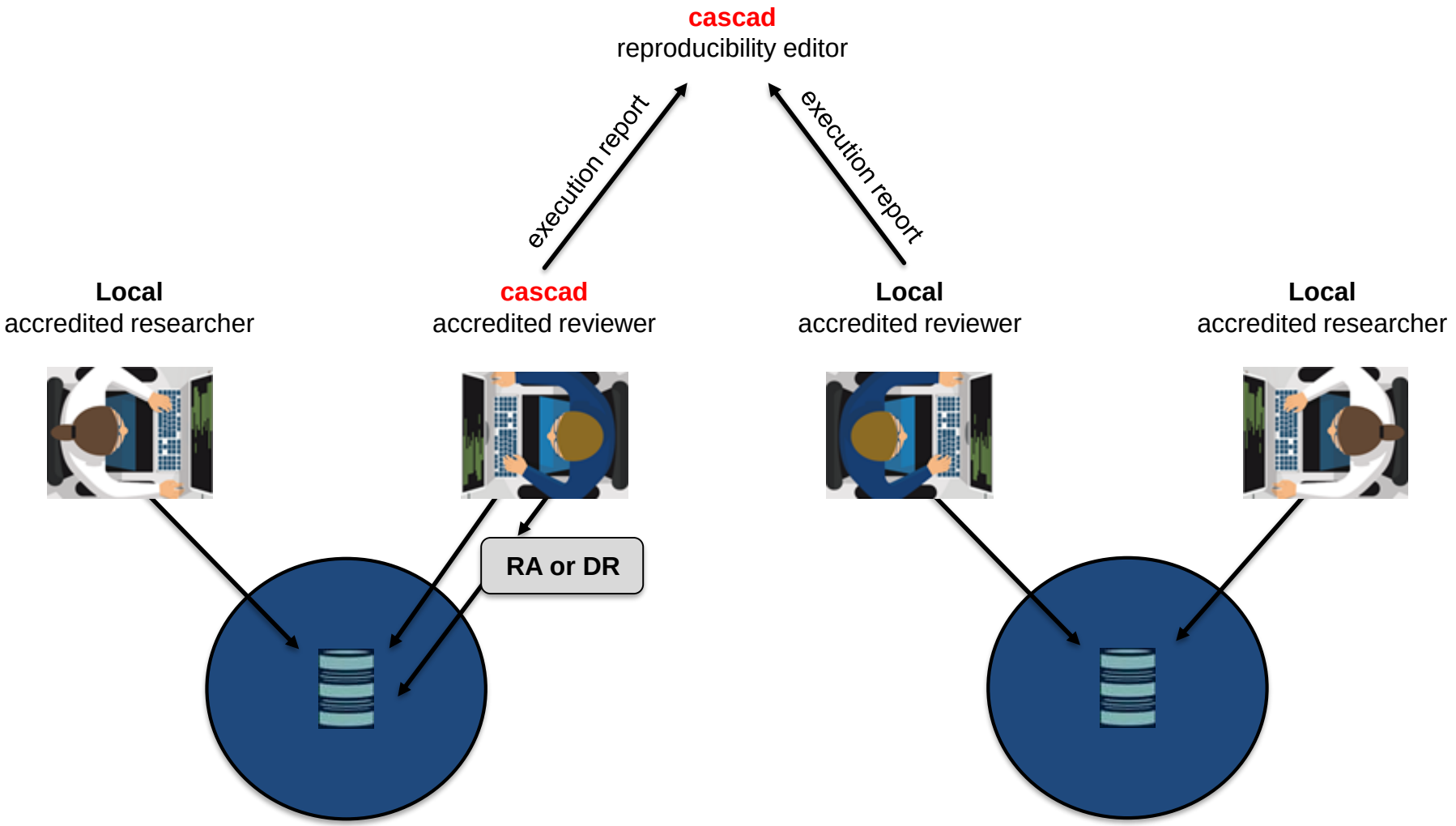


Pérignon, Gadouche, Hurlin, Silberman and Debonnel (2019) Certify reproducibility with confidential data, *Science*

Alternative models for external users



Internal users with direct access



Thank you for your attention !