

Assessing Reproducibility when the Data is Confidential

Lars Vilhuber, Harry Son, Michael Darisse
Cornell University

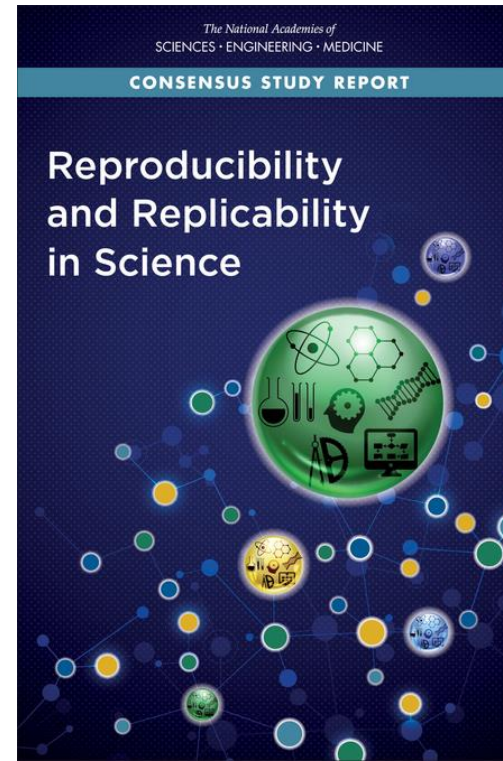
The opinions expressed in this talk are solely the authors, and do not represent the views of the U.S. Census Bureau, the American Economic Association, or any of the funding agencies.





Replication continuum

<https://doi.org/10.17226/25303>



Reproducibility

- Narrow Replication (Pesaran 2003)
- Pure Replication (Hamermesh 2007)
- Verification (Clemens 2015)

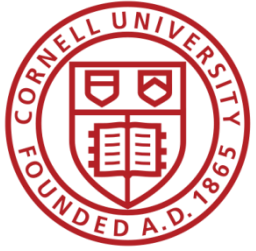
Replicability

- Wide Replication (Pesaran 2003)
- Statistical Replication (Hamermesh 2007)
- Reproduction/Reanalysis (Clemens 2015)

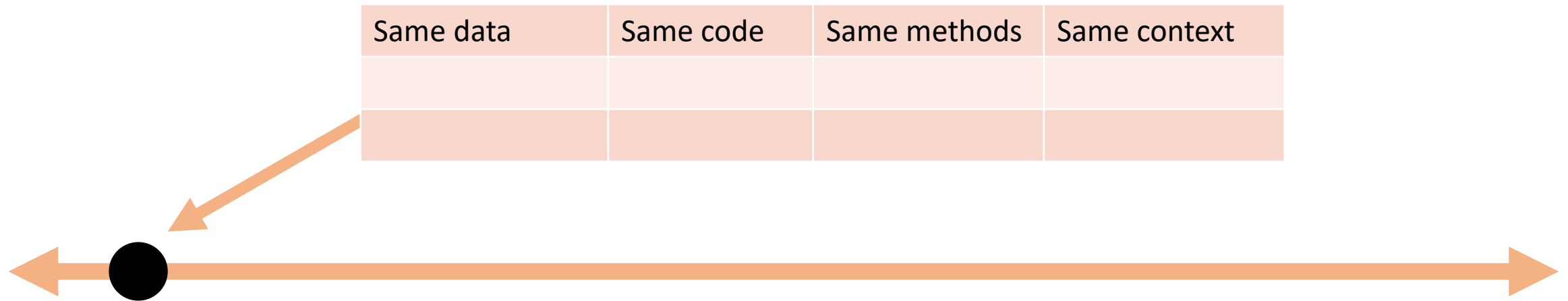
Generalizability

- Wider Replication (Pesaran 2003)
- Scientific Replication (Hamermesh 2007)
- Reanalysis/Robustness (Clemens 2015)





Replication continuum



Reproducibility

- Narrow Replication (Pesaran 2003)
- Pure Replication (Hamermesh 2007)
- Verification (Clemens 2015)





AMERICAN ECONOMIC ASSOCIATION

American Economic Review



The *American Economic Review* is a general-interest economics journal. Established in 1911, the AER is among the nation's oldest and most respected scholarly journals in economics.

American Economic Review: Insights



AER: Insights is designed to be a top-tier, general-interest economics journal publishing papers of the same quality and importance as those in the AER, but devoted to publishing papers with important insights that can be conveyed succinctly.

Journal of Economic Literature



The *Journal of Economic Literature* (JEL), first published in 1969, is designed to help economists keep abreast of and synthesize the vast flow of literature.

Journal of Economic Perspectives



The *Journal of Economic Perspectives* (JEP) fills the gap between the general interest press and academic economics journals.

American Economic Journal: Applied Economics



American Economic Journal: Applied Economics publishes papers covering a range of topics in applied economics, with a focus on empirical microeconomic issues.

American Economic Journal: Economic Policy



American Economic Journal: Economic Policy publishes papers covering a range of topics, the common theme being the role of economic policy in economic outcomes.

American Economic Journal: Macroeconomics



American Economic Journal: Macroeconomics focuses on studies of aggregate fluctuations and growth, and the role of policy in that context.

American Economic Journal: Microeconomics



American Economic Journal: Microeconomics publishes papers focusing on microeconomic theory; industrial organization; and the microeconomic aspects of international trade, political economy, and finance.





AEA Data & Code Availability Policy (2019)

- Access to the data and code must be clearly and precisely documented and is non-exclusive to the authors.
- Authors ... must **provide**, **prior to acceptance**, the data, programs, and other details of the computations **sufficient to permit replication**, as well as **information about access to data and programs**.
- **We test that.**





Current efforts at the AEA

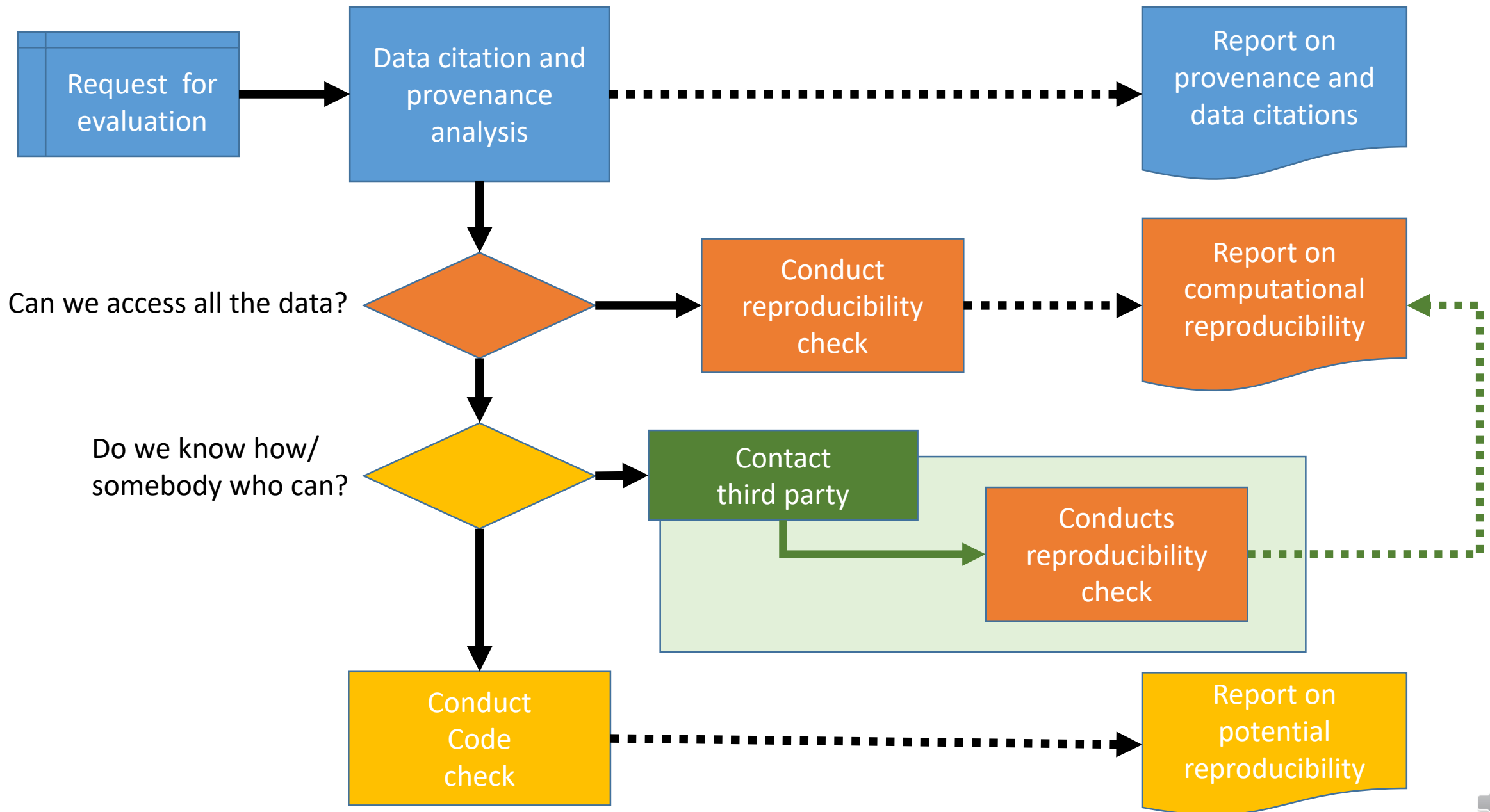
Pre-emptively improve code archives

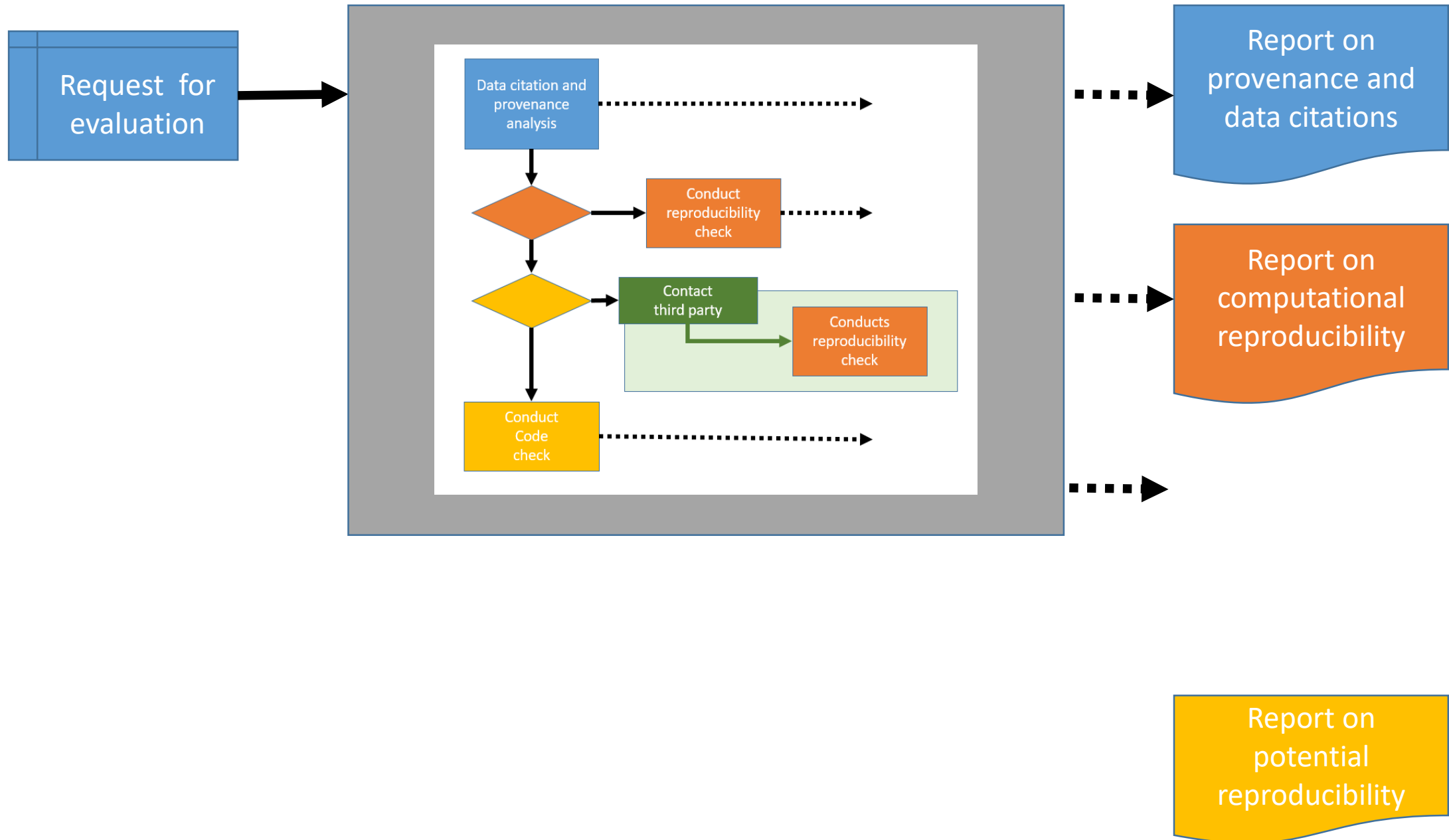
- By conducting reproducibility checks when we can
 - Request data when not freely downloadable
- By working with groups or people that conduct reproducibility checks when we cannot

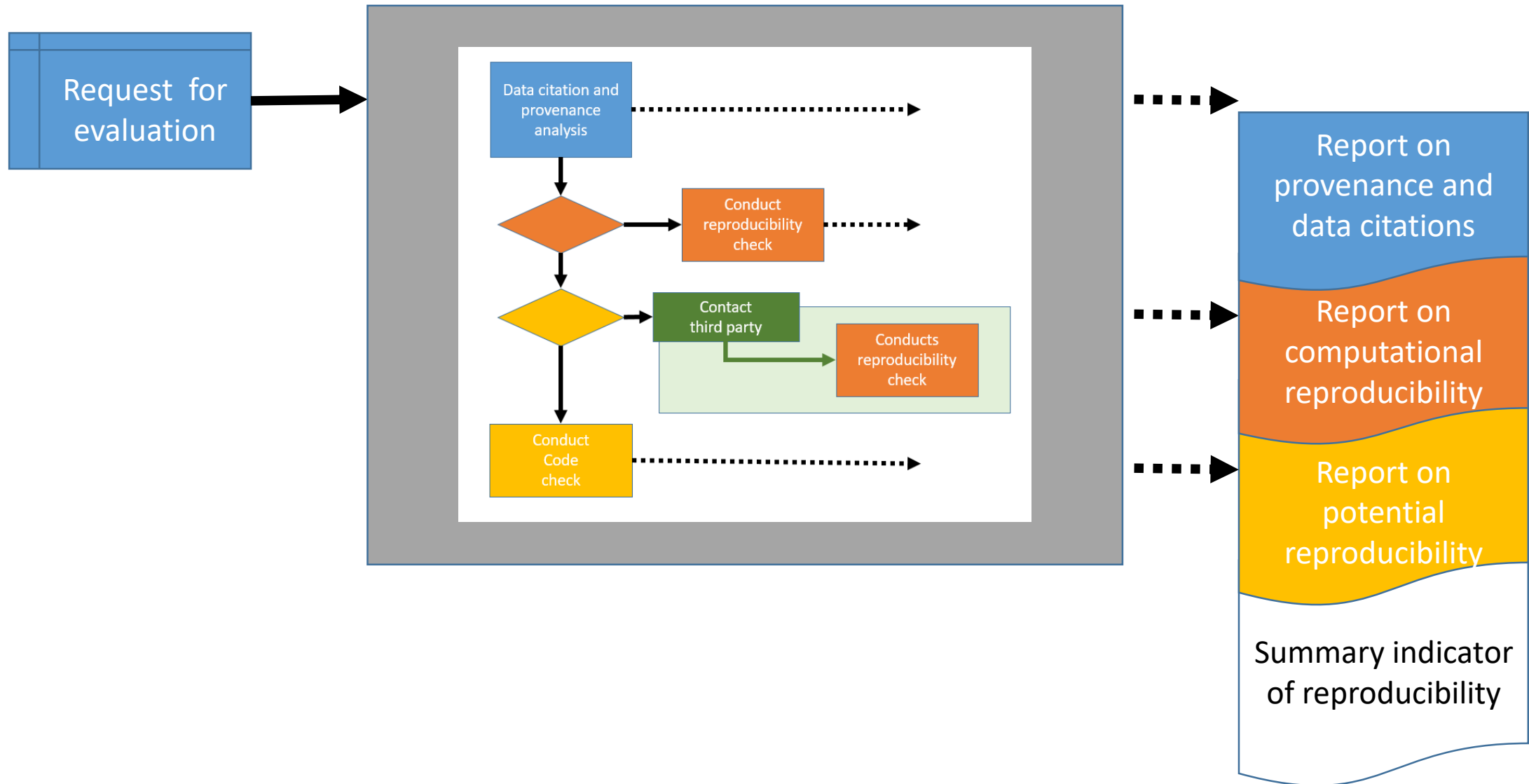
Proprietary and confidential data pose a challenge

How do you check code when the data access is complex?



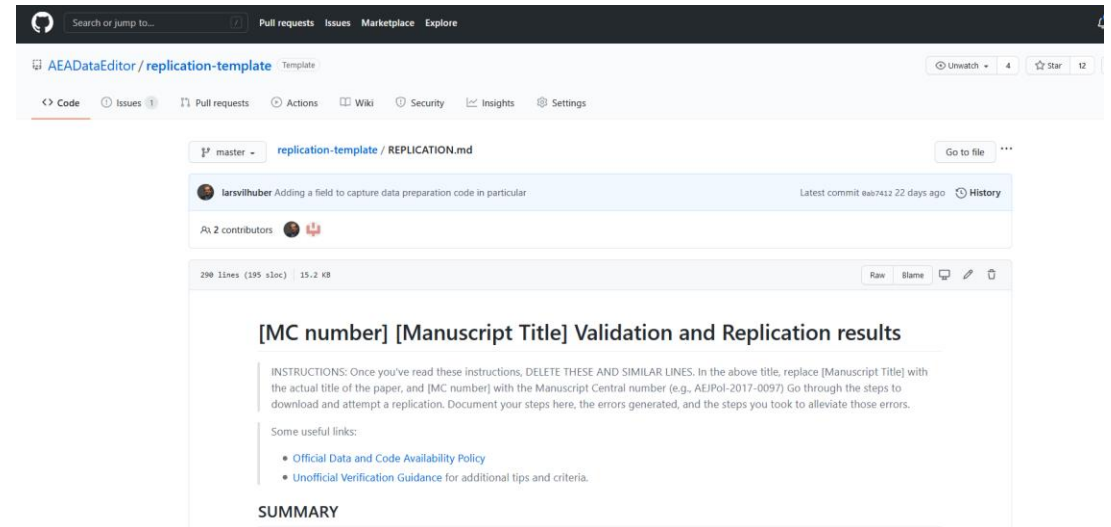
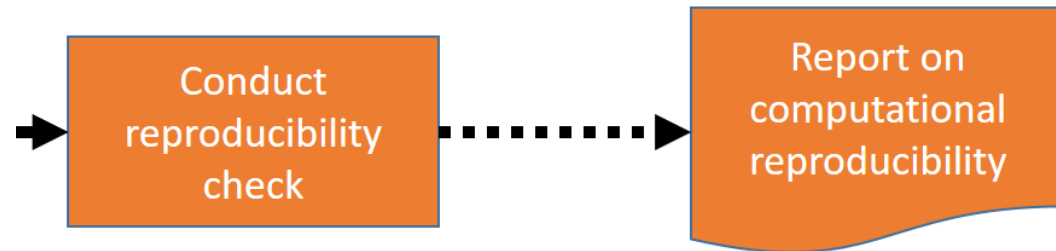






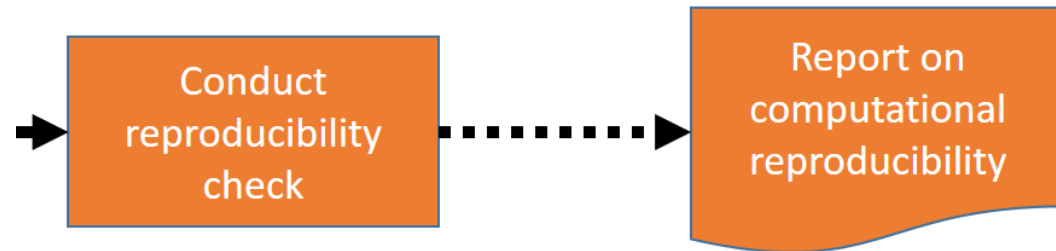


What is the reproducibility check?





What is the reproducibility check?

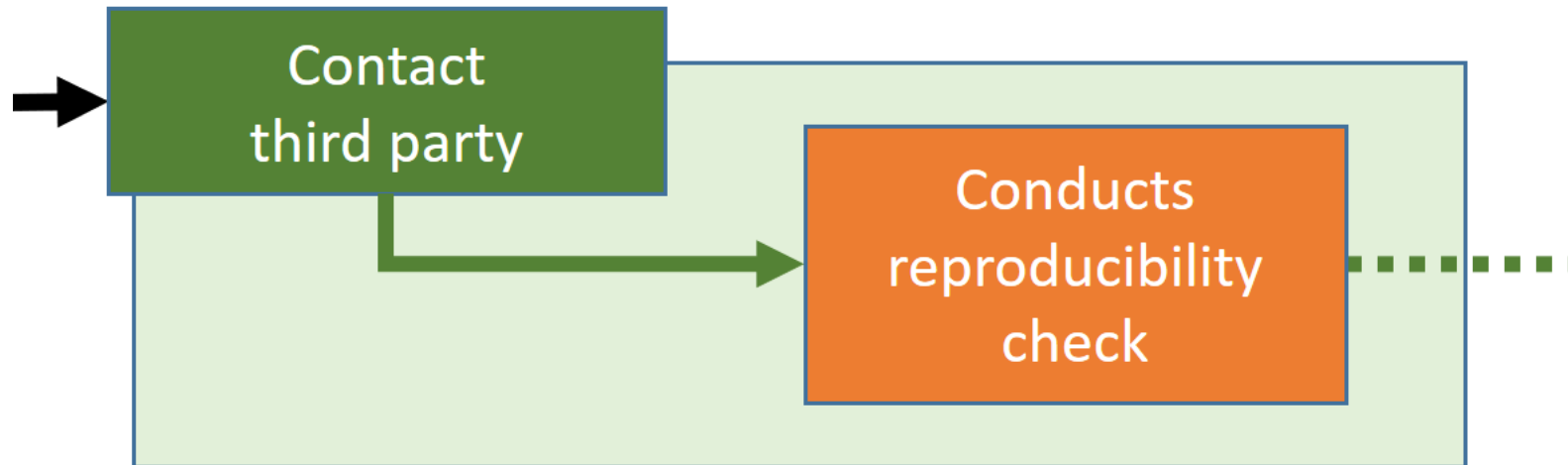


- Data checks
- Code description
- Requirements
 - As stated by author
 - As encountered by replicator
- Verbose description of steps to replicate
- Findings
 - Compare tables
 - Compare figures
 - Compare in-text numbers





Who do we contact when we cannot access the data ourselves?





Who are the 3rd parties?

- cascadi – Using confidential data!
 - DADS
 - French customs data
- Various contributors with access to confidential data
 - Authors
 - Graduate assistants
 - Their institutions
 - Network of known groups with access (less frequent)



A cascadi certification allows researchers to signal the reproducibility nature of their research to their peers





Who are the 3rd parties?

- cascada – Using confidential data!
 - DADS
 - French customs data
- Various contributors with access to confidential data
 - Authors
 - Graduate assistants
 - Their institutions
 - Network of known groups with access (less frequent)

Some other examples:

- IFAU (Swedish data)
- Fed Board (proprietary data)
- Upjohn Institute (US admin data)
- JCT (US tax data)
- Banco do Portugal (LEED)
- IAB (German data)
- *Graduate students* at Wharton, Chicago, Yale, Wisconsin, Cornell





How to prepare the replication package

- README

- Now ask an [RA/ colleague/](#)

AEA Data and Code Guidance



Guidance for authors of data and code supplements and for replicators.

AMERICAN
ECONOMIC
ASSOCIATION

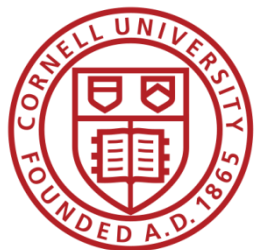
That's our Protocol!

Steps for the Third-party Replicator

- Download the author's replication archive(s) from the designated URL (public, or privately shared)
- Ensure access to any confidential files that are described in the replication archive's README
 - The replicator should consider whether a third-party person not familiar with the author's environment could reasonably rely solely on the instructions in the
- Follow the [checklist](#) to conduct the reproducibility exercise, relying exclusively on the README for instructions and guidance.
- Write a [report](#)
- Send the report to the AEA Data Editor
- Report any interactions with the author in the course of conducting the reproducibility exercise (help, assistance, clarifications)







Social science “guild”



[https://
social-science-
data-editors.
github.io/
guidance/](https://social-science-data-editors.github.io/guidance/)





Some resources

- <https://social-science-data-editors.github.io/guidance/>
- <https://aeadataeditor.github.io/aea-de-guidance/>
 - **template README**
 - discussion of licensing
 - data citation guidance
- German example:
 - Establishment History Panel (BHP) DOI: [10.5164/IAB.BHP7516.de.en.v1](https://doi.org/10.5164/IAB.BHP7516.de.en.v1)
- French verification service “cascad” within French RDC CASD
 - <https://www.casd.eu/en/le-centre-dacces-securise-aux-donnees-casd/certification-de-resultats-cascad-casd/>

