How to Teach Reproducibility in the Classroom

TIER Spring Symposium

Fernando Hoces de la Guardia, BITSS
May 2021 | slides

BITSS

The Berkeley Initiative for Transparency in the Social Sciences works to improve the credibility of science by advancing transparency, reproducibility, rigor, and ethics in research and policy analysis.

Core ACRe

Aleks Bogdanoski

Albert Chae

Fernando Hoces de la Guardia

Katie Hoeberling

Edward Miguel

Lars Vilhuber

We are part of the Center for Effective Global Action (CEGA).

BERKELEY INITIATIVE FOR TRANSPARENCY IN THE SOCIAL SCIENCES

Many Others

CEGA staff

Undergrad and Graduate RAs

Catalysts

Outside Collaborators (Researchers, Programmers)

Table of Contents

1. **Motivation**

2. **A**ccelerating **C**omputational **Re**producibility: Framework

3. ACRe Resources for Teaching Reproducibility in the Classroom: Demo

Motivation 1: Low Replicability and Reproducibility

Replication in Social Sciences (same method, different sample)	Reproduction in Economics (same data and methods)
OSC (2015): 30%-60%	Chang & Li (2015): 43%
Camerer et. al. (2016): ~60%	Gertler et. al. (2017): 14%
Nosek & Camerer et. al. (2018): ~60%	Kingi et. al. (2018): 43%
Klein et. al. (2018): 50%	Wood et. al. (2018): 25%

Motivation 2: More Inclusive Concept Scholarly Output

In addition to reproducibility problems, there are two additional motivations for this projects:

> Clarebout Principle:
>
> "An article about computational science in a scientific publication is not the scholarship itself, it's merely scholarship advertisement. The actual scholarship is the complete software development environment and the complete set of instructions which generated the figures."
>
> *Buckheit and D.L. Donoho (1995, 2009)*

Motivation 2: More Inclusive Concept Scholarly Output

In addition to reproducibility problems, there are two additional motivations for this projects:

> Clarebout Principle:
>
> "An **article** about computational science in a scientific publication is not the scholarship itself, **it's merely scholarship advertisement**. The actual **scholarship is the complete software development environment and the complete set of instructions which generated the figures**."
>
> <div align="right">Buckheit and D.L. Donoho (1995, 2009)</div>

Well discussed potential positive effects on:

- Pedagogy
- Incremental generation of knowledge

Under discussed:

- Possible positive effect on diversity, equity and inclusion: no connections or language skills ("appropriate politeness") required to obtain materials

Motivation 3: Prevent Loss of Knowledge

Every semester, graduate students around **the world** take an Empirical/Applied […] Economics course. A typical assignment consists of reproducing the results of a paper and, possibly, testing the robustness of its results.

Stage	New Knowledge
Scope (select and verify)	Data and code exist?
Assess	Degree of reproducibility for specific part of the paper
Improve	E.g. fixed paths, libraries, added missing files, etc.
Test robustness	Results are robust to additional specifications

Table of Contents

Accelerating Computational Reproducibility: Framework

Context for ACRe

- American Economics Association (AEA) creates first data policy in 2006.

    - Must publish some data (waivers available)

- AEA updates policy in 2019.

    - Must post all data and code. Publication is conditional on verifying reproducibility (if confidential: must document extensively)
    - A new requirement is to post all cleaning code, even for data that is not public
    - See Lars Vilhuber's presentation after this one (same zoom channel) for more information
- Similar updates in Political Science (see AJPS new policy)
- We should expect high levels of computational reproducibility after 2019 (AEA).
- We should not demand 100% reproducibility before, but we could identify the gaps and try to improve some.

Beyond Binary Judgments

Reproductions can easily gravitate towards adversarial exchanges.

- Early career researchers (ECR) have incentives to emphasize unsuccessful reproductions
- Original authors have a more senior position and can use it to deter in-depth reproductions from ECRs.
- The media also focuses on eye-catching headlines

Our approach:

We do not want to say

> "Paper X is (ir)reproducible"

We do want to say

> "Result Y in paper X has a high/low **level** of reproducibility according to **several** reproduction attempts. Moreover, **improvements** have been made to the original reproduction package, **increasing** its reproducibility to a higher level"
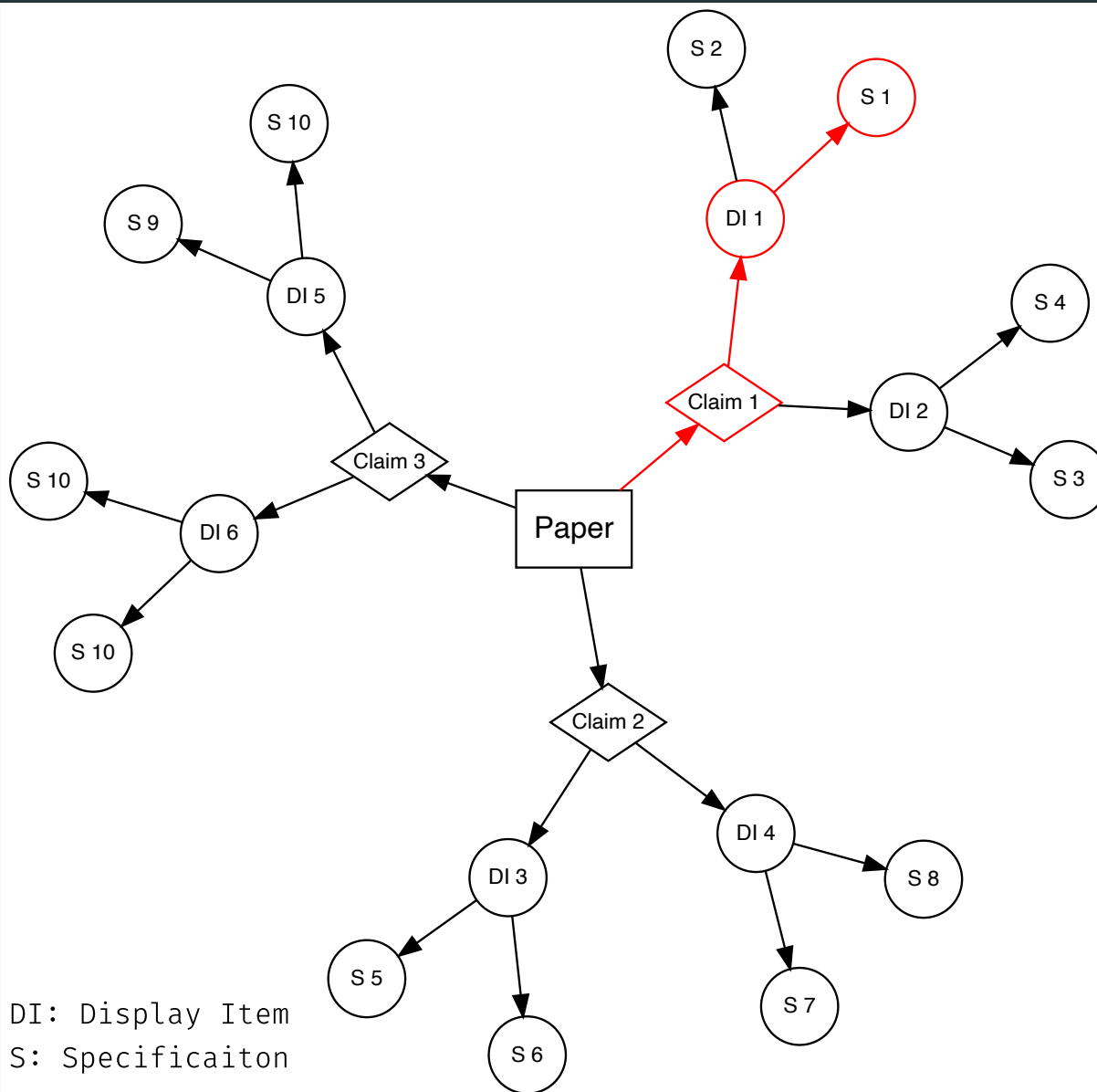
ACRe Framework

Each **reproduction attempt** is centered around scientific **claims**

One paper can contain several claims.

Each claim may be supported by various **display items**: tables, figures & inline results.

A reproduction attempt is at the claim level, and reproducers must record their **specifications** of interest.

DI: Display Item
S: Specificaiton

Large part of this exercise is about standardization

- **Computational Reproduction (or Reproduction)**

- **Replication (will not mention this term again!)**

- Reproduction attempt

- **Reproduction package**

- Claim

- Display item

- Specification

- Preferred specification

- **Raw data**

- **Analysis data**

- **Candidate paper**

- **Declared paper**

- Reproduction tree

- Complete Workflow

- Computationally Reproducible from Analytic data (CRA)

- Computationally Reproducible from Raw data (CRR)

- Reasonable test
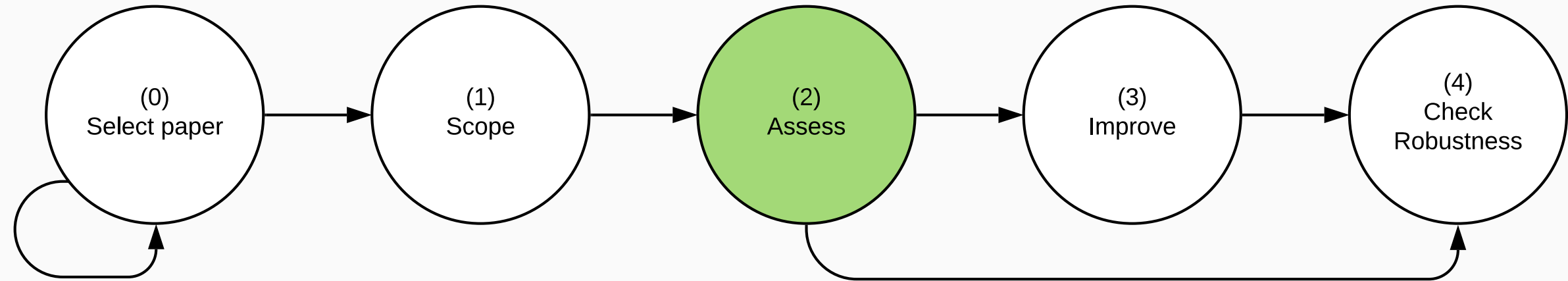
- Feasible test

- Minimal effort

Scope
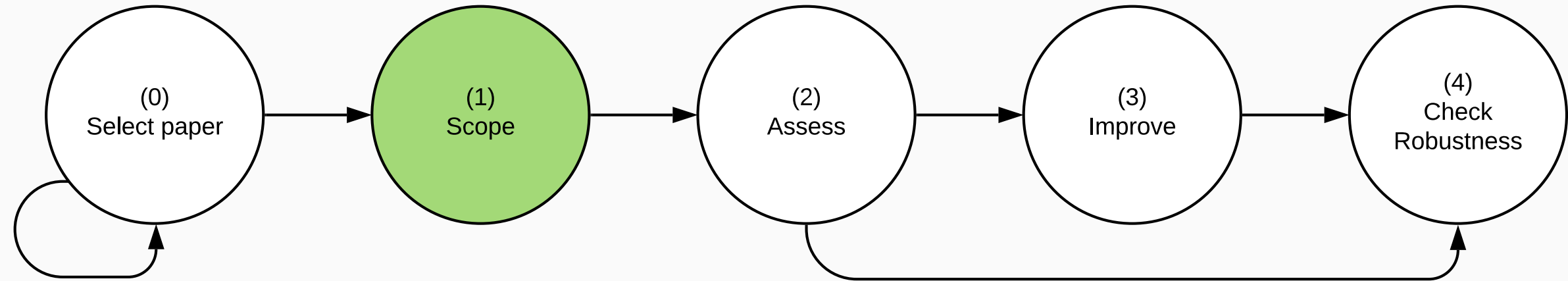
Assess

Two main parts for assessment:

1. Find all the elements behind a display item
2. Score the reproducibility of that display item

Identify All the Elements Behind a Display Item

Reproducers will be asked to draw a clear connection to the raw data sources mentioned in the paper and the display item under reproduction.

Data sources

Connect the data sources in the paper's text with specific raw data files.

Analytic data sets

Describe each analytic data file.

Code files

Inspect all code files and record all their inputs and outputs.

With all the information recorded above, reproducers can use the **ACRe Diagram Builder** to generate a **reproduction tree**.

ReproducibiliTREE

```
table1.tex
    |___[code] analysis.R
        |___analysis_data.dta
            |___[code] final_merge.do
                |___cleaned_1_2.dta
                |    |___[code] clean_merged_1_2.do
                |        |___merged_1_2.dta
                |            |___[code] merge_1_2.do
                |                |___cleaned_1.dta
                |                |    |___[code] clean_raw_1.py
                |                |        |___raw_1.dta
                |                |___cleaned_2.dta
                |                    |___[code] clean_raw_2.py
                |                        |___raw_2.dta
                |___cleaned_3_4.dta
                    |___[code] clean_merged_3_4.do
                        |___merged_3_4.dta
                            |___[code] merge_3_4.do
                                |___cleaned_3.dta
                                |    |___[code] clean_raw_3.py
                                |        |___raw_3.dta
                                |___cleaned_4.dta
                                    |___[code] clean_raw_4.py
                                        |___raw_4.dta
```

```
              Levels of Computational Reproducibility
                 with Proprietary/Confidential Data
              (P denotes "partial", C denotes "complete")
                        | Availability of materials, and reproducibility |
                        |------------------------------------------------|
                        |         | Instr. |     |         | Instr.|     |
                        |Analysis| Analysis|     | Cleaning| Raw   |     |
                        |Code    | Data    | CRA | Code    | Data  | CRR |
                        | P | C  | P  | C  |     | P  | C  | P | C |     |
                        ---------|---------|-----|---------|-------|-----|
L1: No materials.................| -   -  | -   -  | -  | -   -  | -   - | -  |
--------------------------------|--------|--------|-----|--------|-------|-----|
L2: Only code ..................| ✔   ✔  | -   -  | -  | -   -  | -   - | -  |
L3*: Partial analysis data & code| ✔   ✔  | ✔   -  | -  | -   -  | -   - | -  |
L4*: All analysis data & code....| ✔   ✔  | ✔   ✔  | -  | -   -  | -   - | -  |
L5*: Proof of third party CRA....| ✔   ✔  | ✔   ✔  | ✔  | -   -  | -   - | -  |
--------------------------------|--------|--------|-----|--------|-------|-----|
L6: All cleaning code............| ✔   ✔  | ✔   ✔  | -  | ✔   ✔  | -   - | -  |
L7: Some instr. for raw data.....| ✔   ✔  | ✔   ✔  | -  | ✔   ✔  | ✔   - | -  |
L8*: All instr. for raw data.....| ✔   ✔  | ✔   ✔  | -  | ✔   ✔  | ✔   ✔ | -  |
L9*: L8 + Proof of CRA...........| ✔   ✔  | ✔   ✔  | ✔  | ✔   ✔  | ✔   ✔ | -  |
L10*:Proof of third party CRR....| ✔   ✔  | ✔   ✔  | ✔  | ✔   ✔  | ✔   ✔ | ✔  |
```
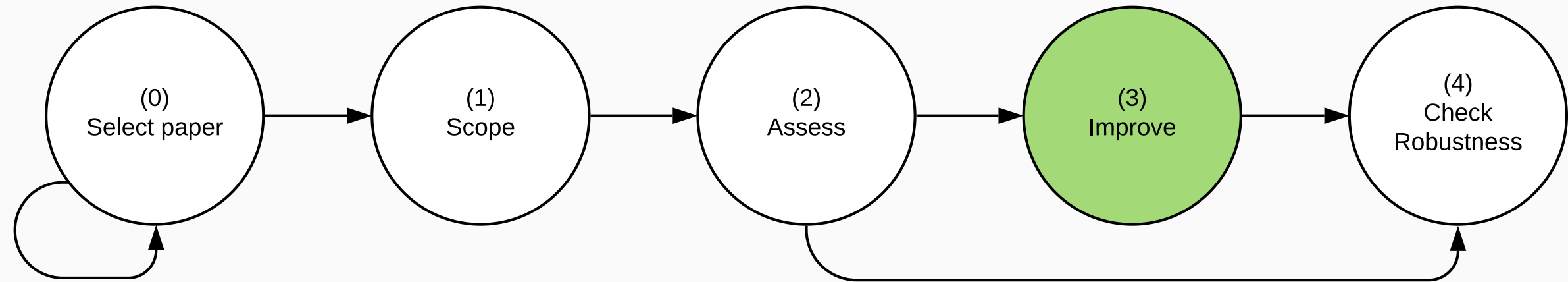
Improvements
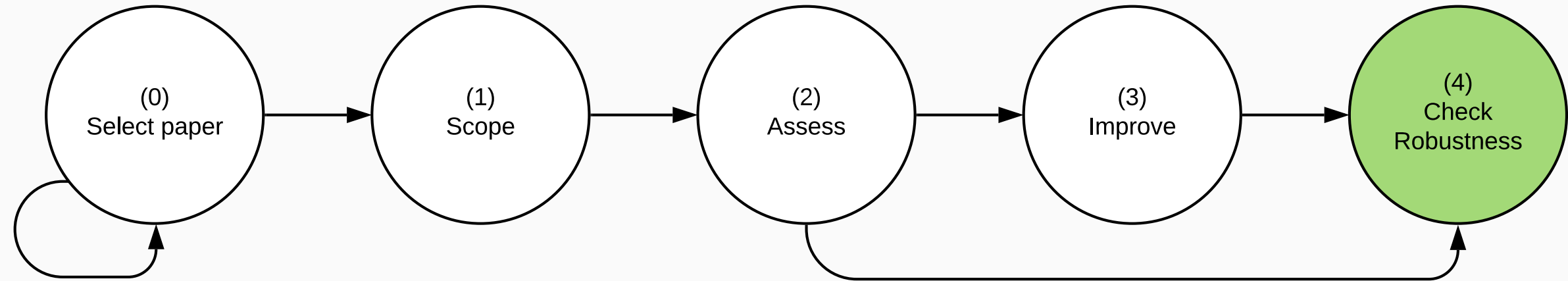
Three types of improvements:

1. Improvements at the paper level
2. Improvements at the display-item level
3. Specific future improvements

Robustness Checks

Two main parts for robustness:

1. Increase the number of robustness checks
2. Justify the appropriateness of a specific test

Robustness

Robustness checks: any possible change in a computational choice, both in data analysis and data cleaning

Reasonable specifications (Simonsohn et. al., 2018):

1. Sensible tests of the research question
2. Expected to be statistically valid, and
3. Not redundant with other specifications in the set.

Reproducers will be able to record two types of contributions:

- Mapping the universe of robustness checks
- Proposing a specific robustness check
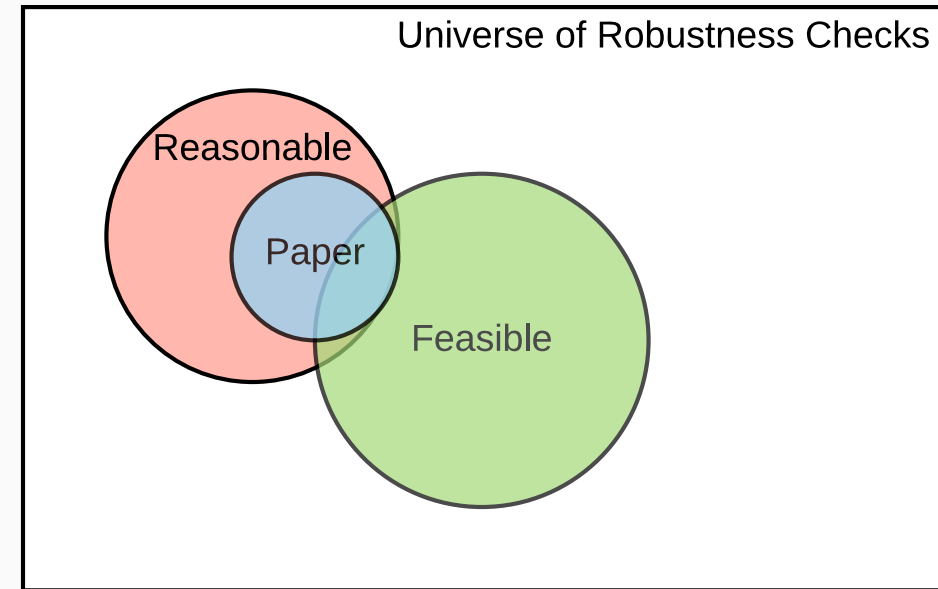
Table of Contents

Social Science Reproduction Platform (SSRP): Demo

www.socialsciencereproduction.org

Key features:

- Record a reproduction (stages, assessment, improvements, robustness)

- Forum

- ACRe Guide

- Strong emphasis on constructive exchanges

- Coming soon: Reproducibility Dashboard, and many other features

Target users:

- Students in graduate empirical courses

- Undergraduates thesis projects

- Researchers conducting reproductions

- Researchers conducting original research (for internal purposes)

Example 1: There is no reproduction package

Subject: Reproduction package for [`"Title of the paper"`]

Dear Dr. [`Lastname of Corresponding Author`],

I am contacting you to request a reproduction package for your paper titled [`Title`] which was published in [`Reference`]. A reproduction package may contain (raw and/or analytic) data, code, and other documentation that makes it possible to reproduce paper. Would you be able to share any of these items?

I am a [`position`] at [`Institution`], and I would like to reproduce the results, tables, and other figures using the reproduction materials mentioned above. I have chosen this paper because [`add context ... `]. **Unfortunately, I was not able to locate any of these materials on the journal website, Dataverse** [`or other data and code repositories`]**, or in your website**.

I will record the result of my reproduction attempt on ACRe [...]. With your permission, I will also record the materials you share with me, which would allow access for other reproducers and avoid repeated requests directed to you. Please let me know if there are any legal or ethical restrictions that apply to all or parts of the reproduction materials so that I can take that into consideration during this exercise.

In addition to your response above, would you be available to respond to future (non-repetitive) inquiries from me or other reproducers conducting an ACRe excercise? **Though your cooperation with my and/or any future request would be extremely helpful, please note that you are *not required to comply*.**

Since I am required to complete this project by [`date`], I would appreciate your response by [`deadline`].

Let me know if you have any questions. Please also feel free to contact my supervisor/instructor [`Name (email)`] for further details on this exercise. Thank you in advance for your help!

Example 1: Following up on additional materials

Template email:

> **Subject:** Clarification for reproduction materials for `["Title of the paper"]`
>
> Dear Dr. `[Lastname of Corresponding Author]`,
>
> Thank you for sharing the materials. They have been immensely helpful for my work.
>
> Unfortunately, I ran into a few issues as I delved into the reproduction exercise, and I think your guidance would be helpful in resolving them. **`[Describe the issues and how you have tried to resolve them. Describe whatever files or parts of the data or code are missing. Refer to examples 1 and 2 below for more details]`**.
>
> Thank you in advance for your help.
>
> Best regards,
> `[Reproducer]`

An example of well described issues:

> Specifically, I am attempting to reproduce OUTPUT X (e.g., table 1, figure 3). I found that the following components are required to reproduce to reproduce OUTPUT X:

```
OUTPUT X
    └──[code] formatting_table1.R
    ├──output1_part1.txt
    │    └──[code] output_table1.do
    │        └──[data] analysis_data01.csv
    │            └──[code] data_cleaning01.R*
    │                └──[data] UNKNOWN
    └──output1_part2.txt
        └──[code] output_table2.do
            └──[data] analysis_data02.csv
                └──[code] data_cleaning02.R
                    └──[data] admin_01raw.csv*
```

> I have marked with an asterisk (*) the items that I could not find in the reproduction materials: **data_cleaning01.R** and **admin_01raw.csv**. After accessing these files, I will also be able to identify the name of the raw data set required to obtain output1_part1.txt. This is to let you know that I may need to contact you again if I cannot find this file (labeled as **UNKNOWN** above) in the reproduction materials.
>
> I understand that this request will require some work for you or somebody in your research group, but I want to assure you that I will add these missing files to the reproduction package for your paper on the ACRe platform. **Doing this will ensure that you will not be asked twice for the same missing file.**

Ok, I get it. But what is in for me?

- Standardized homework/project: everything is set up in terms of structure and deliverables.

- Easy to grade (homework format).

- Easy to guide and oversee (undergraduate dissertation format).

- Easy to setup as an independent study.

- Reduces duplication of requests to authors.

- Facilitates a constructive exchange of ideas.

    - When emailing authors.
    - When discussion reproduction attempts.

- Personal satisfaction that you're contributing a public good to the profession!

Ok, I Am Interested. What's Next?

Check the platform and materials yourself www.socialsciencereproduction.org
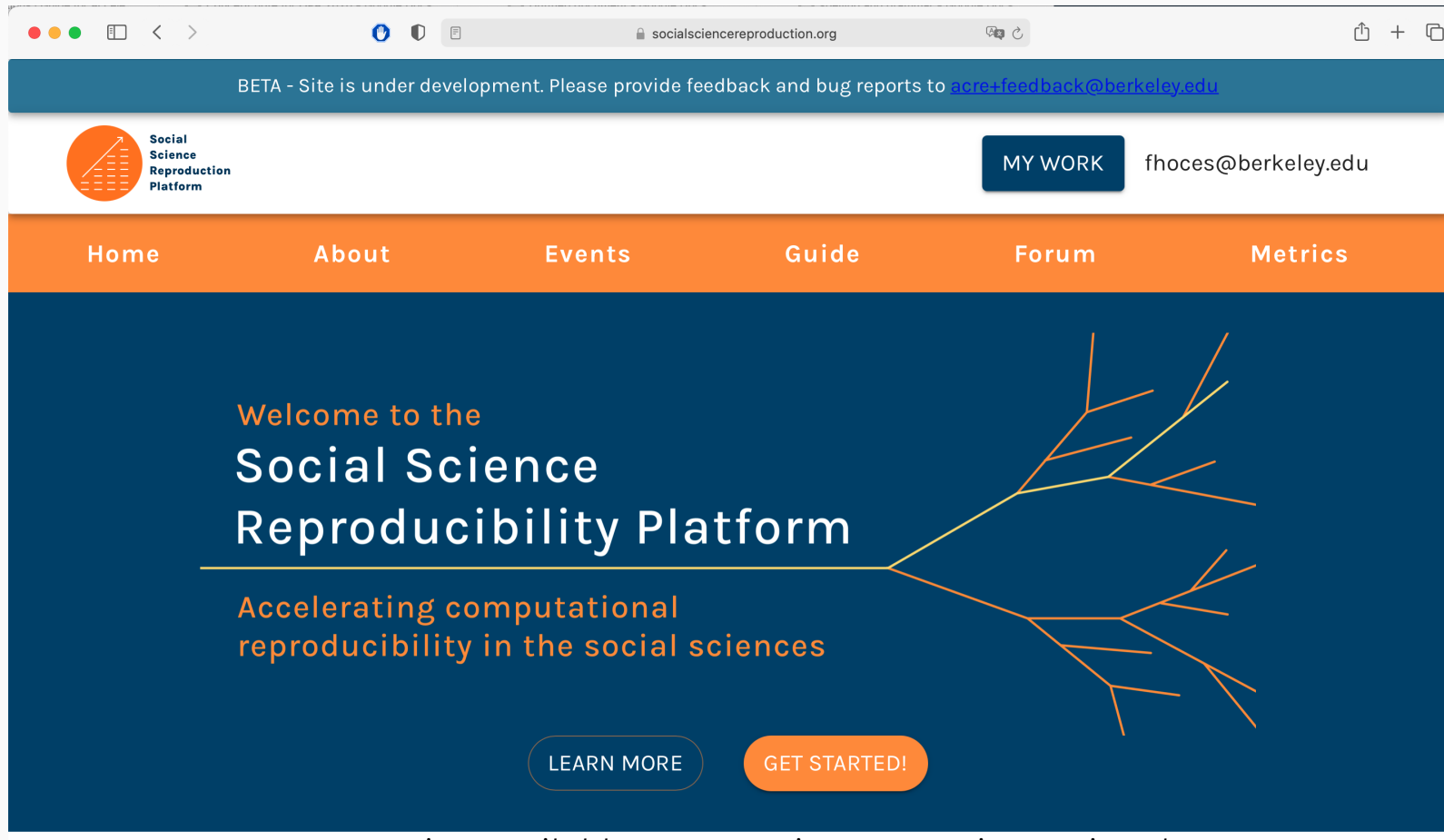
Beta version available now. Version 1.0 coming up in July!

Contact us at acre@berkeley.edu if you are interested in using the platform as part of a course.