Data Access and Research Transparency
Why Does Openness in Political Science Matter?

Valentina González Rostani

Ph.D. Student, Department of Political Science
University of Pittsburgh

May 3, 2021

Outline

- Political science debate

- New ethics guidelines

- How do we benefits from increased openness?

- How are we doing?

- Tools, workflows and practical guidance

- Good examples: building community

Political Science Debate

Political Science Debate

- 1995 "*Replication, Replication*" by King (1995) states the importance of sharing data and replication files.
- 2003 Leading journals requested replication files (JPR, ISQ, II, JCR)
- 2012 American Political Science Association (APSA) new **Guide to Professional Ethics in Political Science** (see guide).
- 2014 27 leading political science journals committed to greater data access described in the **Journal Editors Transparency Statement** (JETS)
- 2015 Petition to delay the implementation of the new guidelines (see link)

New Ethics Guidelines

New Ethics Guidelines

APSA ethics guidlines suggests: *"researchers have an ethical obligation to facilitate the evaluation of their evidence based knowledge claims through* **data access, production transparency, and analytic transparency** *so that their work can be tested or replicated"*

Data Access
Reference the data used, and if it is generated by researchers then they should provide access to those data or explain why they cannot.

Producing Transparency
Should offer a full account of the procedures used to collect or generate the data.

Analytic Transparency
Evidence-based knowledge claims should provide a full account of how they draw their analytic conclusions from the data (example: codes)

Specific guidelines

1. Make all data available
2. Delineate clearly the analytical procedures used to generate the data
3. Delineate clearly the analytical procedures used to analyze the data

How Do We Benefits?

How do we benefits from increased openness?

1. **Protect scholarly integrity**: avoid p-hacking and data fabrication Example: Diederik Stapel *"Nobody ever checked my work. They trusted me... I did everything myself, and next to me was a big jar of cookies. No mother, no lock, not even a lid... Every day, I would be working and there would be this big jar of cookies, filled with sweets within reach, right next to me -with nobody ever near. All I had to do was take it."* (quoted in Borsboom and Wagenmakersm, 2013)

2. **Evaluate claims**: avoid honest mistakes. Allow researchers' audience to evaluate claims and its logic as valid (Lupia & Elman, 2014).

3. **Community building knowledge**: learn from how other social scientists generate their conclusions. Shared data$\rightarrow$ valuable public good.

How Are We Doing?

How Are We Doing?

- Key (2016) shows 3 of JETS signatories journals –*International Organizations, American Journal of Political Research, and Political Analysis*– and replication materials were available for more than 80%

- Stockemer, Koehler, and Lentz (2018) look at three journals –*Electoral Studies, Party Politics, and Journal of Elections, Public Opinion Parties*– in 2015 and they find

  Regarding the data access
  - Only a few articles have data accesible online.
  - Once they requested around 50% of the authors make available the data

  Regarding replicability, from those articles with data available
  - In 70% of the cases they confirm the results
  - In 25% of cases, replication is impossible due to poor organization of the data and/ or code
  - In 5% of articles, the replication results are different

Tools, Workflows and Practical Guidance

Tools

- Pre-registration: pre-analysis plans
- Methodological Appendices: how an author collected, generated, and analyzed data can take multiple forms (Kapiszewski & Karcher, 2019)
- Personal websites
- Digital repositetories Github, Harvard Dataverse, OSF, ICPSR, UK Data Service, CISER.
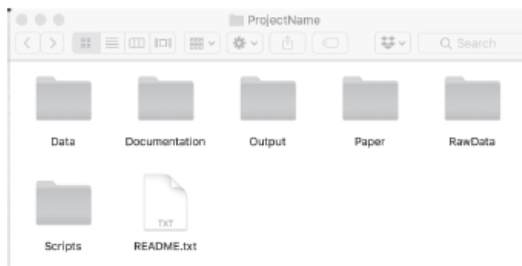
Workflow (Christensen, Freese, & Miguel, 2019)

Practical Guidance

- Create an account in these **digital repositories**
- **Be careful when managing databases**. For example:
    - Always have the raw data in a directory, and the cleaned in another one.
    - Clean the data using do-file/scripts.
    - Never manipulate the data by hand.
- Commit yourself with good workflow practices. Have a structure of files that you and others can understand, and a "README" file describing the contents.
- Become familiarized with github and other languages that will help you to collaborate and share with the research community. Github will allow you to track changes and recover previous versions.
- Create your **website**, there are many tools (see examples weebly, github, pitt)

Workflow

For more resources here some books' recommendations that are available from the Pitt Library:

- For detailed suggestions for *Reproducible Workflow* see chapter 11 of Christensen et al. (2019)
- For an introduction to Git and GitHub see the book Beginning Git and GitHub

Source (Christensen et al., 2019)

Building Community

Building Community

Two examples of young political scientists who by sharing their packages and data are also providing a valuable public good for the whole society:

- José Incio
    - Covid-Perú (R package) and webpage following fake news, and regional statistics.
    - Escribe a tu congresista - Write to your representative!
    - Moving from Stata to R
- Nicolas Schmidt
    - Electoral indicators (R package) estimate various indicators such as electoral disproportionality, party nationalization, effective number of parties.
    - Speech (R package) parliamentary speeches in Uruguayan
    - Boreluy (R package) Electoral data from Uruguay 1918-2019.

Thank you!

References I

Christensen, G., Freese, J., & Miguel, E. (2019). *Transparent and Reproducible Social Science Research: How to Do Open Science*. Univ of California Press. (Google-Books-ID: PvqVDwAAQBAJ)

Kapiszewski, D., & Karcher, S. (2019, October). Transparency in Practice in Qualitative Research. Retrieved 2020-09-20, from https://preprints.apsanet.org/engage/apsa/article-details/5d94c7762f41c7001256af6d doi: 10.33774/apsa-2019-if2he-v2

Key, E. M. (2016, April). How Are We Doing? Data Access and Replication in Political Science. *PS: Political Science & Politics*, *49*(2), 268–272. Retrieved 2020-09-20, from http://www.cambridge.org/core/journals/ps-political-science-and-politics/article/how-are-we-doing-data-access-and-replication-in-political-science/8EE49C7FE45B46F20A50D9271BB3AAC4 (Publisher: Cambridge University Press) doi: 10.1017/S1049096516000184

References II

King, G. (1995). Replication, Replication. *PS: Political Science and Politics*, *28*(3), 444–452. Retrieved 2020-09-20, from http://www.jstor.org/stable/420301 (Publisher: [American Political Science Association, Cambridge University Press]) doi: 10.2307/420301

Lupia, A., & Elman, C. (2014, January). Openness in Political Science: Data Access and Research Transparency: Introduction. *PS: Political Science & Politics*, *47*(1), 19–42. Retrieved 2020-09-20, from http://www.cambridge.org/core/journals/ps-political-science-and-politics/article/openness-in-political-science-data-access-and-research-transparency/0B189FC1097A4D062E57F805E8F07BD0 (Publisher: Cambridge University Press) doi: 10.1017/S1049096513001716

References III

Stockemer, D., Koehler, S., & Lentz, T. (2018, October). Data Access, Transparency, and Replication: New Insights from the Political Behavior Literature. *PS: Political Science & Politics*, *51*(4), 799–803. Retrieved 2020-09-20, from http://www.cambridge.org/core/journals/ps-political-science-and-politics/article/data-access-transparency-and-replication-new-insights-from-the-political-behavior-literature/64CA07CBA652E299079FF32BC5A6DCB3 (Publisher: Cambridge University Press) doi: 10.1017/S1049096518000926