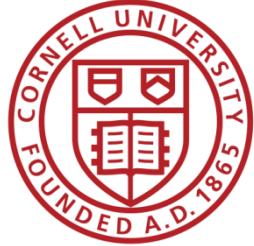




Practices for Data Transparency and Reproducibility

Lars Vilhuber
Cornell University
2020-09-24

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.



On the menu today

- Data provenance: what you want is what you should get
- README when you can
- The Archive at the End of the Universe: better than your average website
 - Subplot: When its confidential?



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.

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.

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: Macroeconomics



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

AMERICAN ECONOMIC ASSOCIATION

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 Perspectives



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

American Economic Journal: Economic Policy

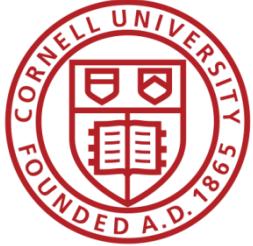


American Economic Journal: Economic Policy publishes papers covering a range of topics, the common theme being the role of economic policy in economic outcomes.

American Economic Journal: Microeconomics

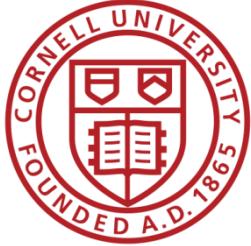


American Economic Journal: Microeconomics publishes papers focusing on microeconomic theory; industrial organization; and the microeconomic aspects of international trade, political economy, and finance.



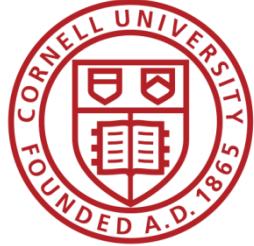
AEA Data & Code Availability Policy (2019)

- It is the policy of the American Economic Association to publish papers only if the data used in the analysis are **clearly and precisely documented** and **access to the data and code is clearly and precisely documented and is non-exclusive to the authors.**
- Authors of accepted papers that contain empirical work, simulations, or experimental work 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**.



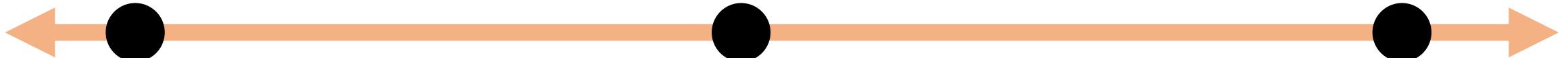
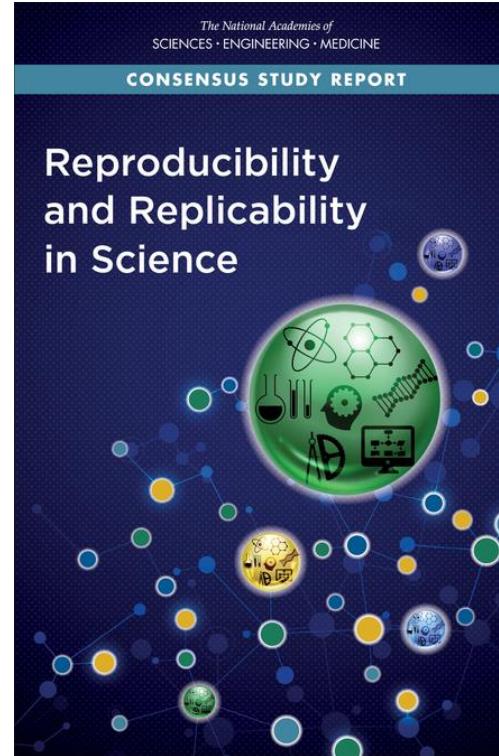
Current efforts at the AEA

- **Pre-emptively improve code archives**
 - By conducting reproducibility checks when we can
 - By working with groups that conduct reproducibility checks when we cannot
- **Better archives**
 - Greater transparency of the code and data archives
- **Better provenance tracking**
 - Leave code where it is when appropriate
 - Leave data where it is almost always
 - Display that information



Replication continuum

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



Reproducibility

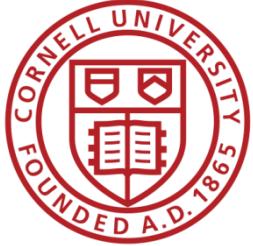
- Narrow Replication (Pesaran 2003)
- Pure Replication (Hamermesh 2007)
- Verification (Clemens 2015)

Replicability

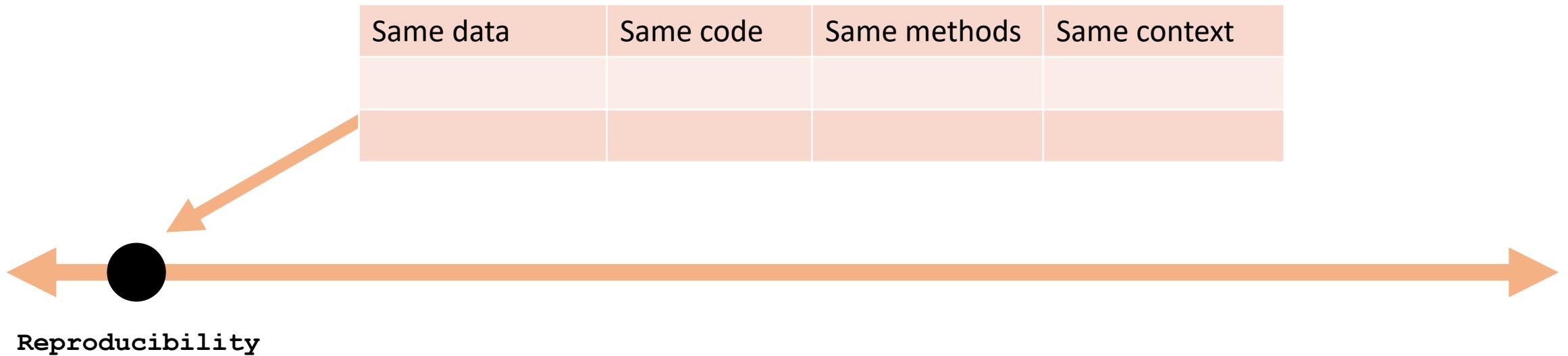
- Wide Replication (Pesaran 2003)
- Statistical Replication (Hamermesh 2007)
- Reproduction/Reanalysis (Clemens 2015)

Generalizability

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

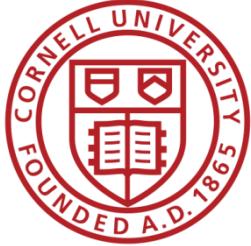


Replication continuum



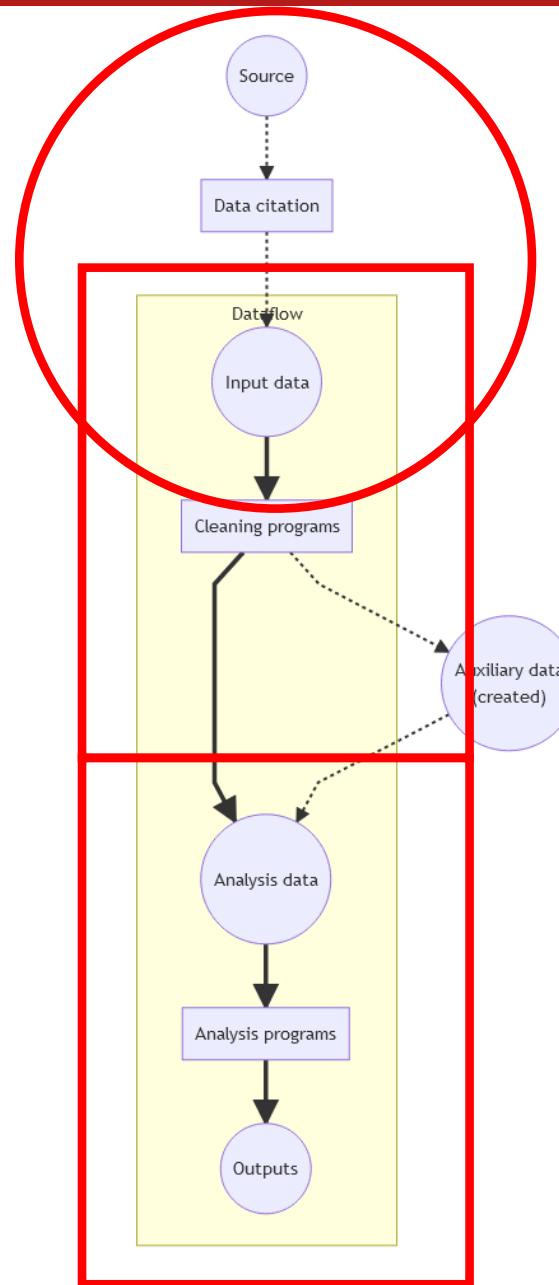
- Narrow Replication (Pesaran 2003)
- Pure Replication (Hamermesh 2007)
- Verification (Clemens 2015)

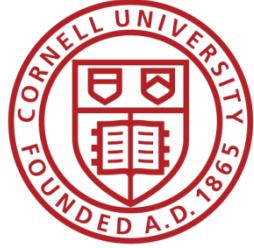
Data provenance



Simplified model

- Basic data-oriented flow
- Historically: analysis data
- More recently: input data, too
- But where does it come from?
("provenance")





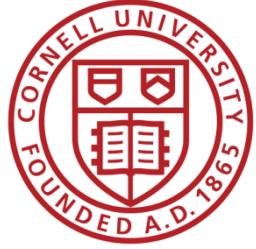
Poor citation practices

- **Macrodata:**

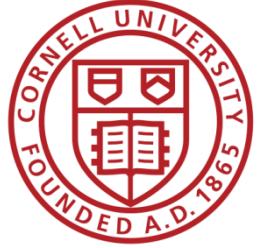
“We use data downloaded from
the Bureau of Economic Analysis...”

- **Microdata:**

“... this paper uses data from
the Current Population Survey...”

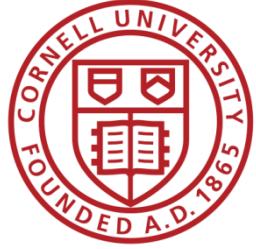


An analogy...



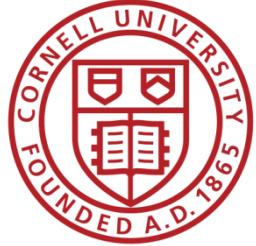
Would you buy a car from this guy?





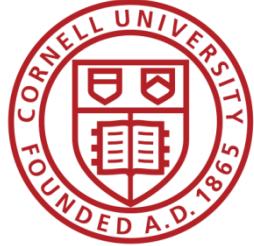
Provenance!

- Does the sales person have a good record?
- Where does the car come from?
- What do we know about the car?



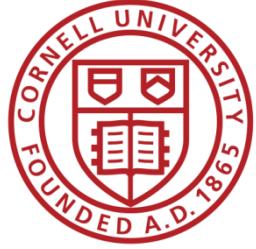
Would you use this data?

```
00000000 cfd0 e011 b1a1 e11a 0000 0000 0000 0000  
00000010 0000 0000 0000 0000 003e 0003 fffe 0009  
00000020 0006 0000 0000 0000 0000 0000 0004 0000  
00000030 008f 0000 0000 0000 1000 0000 fffe ffff  
00000040 0000 0000 fffe ffff 0000 0000 008b 0000  
00000050 008c 0000 008d 0000 008e 0000 ffff ffff  
00000060 ffff ffff ffff ffff ffff ffff ffff ffff  
*  
0000200 0809 0010 0600 0005 209a 07cd c0c9 0000  
0000210 0306 0000 00e1 0002 04b0 00c1 0002 0000  
0000220 00e2 0000 005c 0070 0001 4c00 2020 2020  
0000230 2020 2020 2020 2020 2020 2020 2020 2020  
*  
0000290 2020 2020 2020 2020 0042 0002 04b0 0161  
00002a0 0002 0000 013d 0002 0001 009c 0002 000e  
00002b0 0019 0002 0000 0012 0002 0000 0013 0002  
00002c0 0000 01af 0002 0000 01bc 0002 0000 003d  
00002d0 0012 0000 000f 3f1b 27f6 0038 0000 0000  
00002e0 0001 0258 0040 0002 0000 008d 0002 0000  
00002f0 0022 0002 0000 000e 0002 0001 01b7 0002
```



Or would you trust this data?

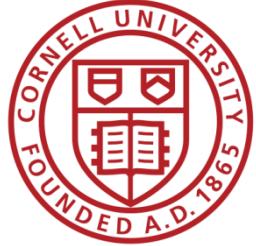




Provenance!

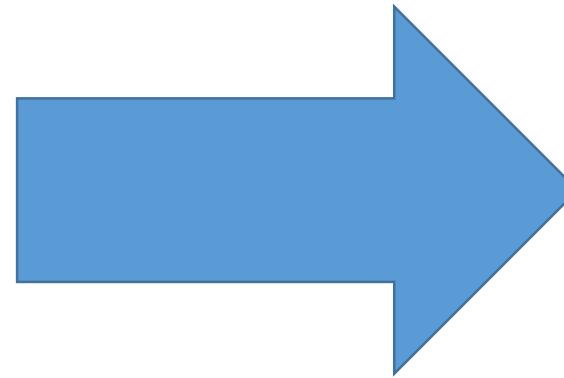
- Does the provider have a good record?
- Where do the data come from?
- What do we know about the data?

Metadata!

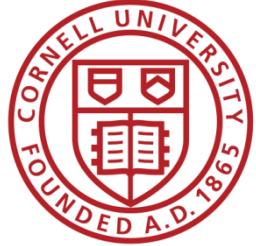


“It’s a file called stockmarket.xlsx”

```
00000000 cfd0 e011 b1a1 e11a 0000 0000 0000 0000  
00000010 0000 0000 0000 0000 003e 0003 fffe 0009  
00000020 0006 0000 0000 0000 0000 0000 0004 0000  
00000030 008f 0000 0000 0000 1000 0000 fffe ffff  
00000040 0000 0000 fffe ffff 0000 0000 008b 0000  
00000050 008c 0000 008d 0000 008e 0000 ffff ffff  
00000060 ffff ffff ffff ffff ffff ffff ffff ffff  
*  
0000200 0809 0010 0600 0005 209a 07cd c0c9 0000  
0000210 0306 0000 00e1 0002 04b0 00c1 0002 0000  
0000220 00e2 0000 005c 0070 0001 4c00 2020 2020  
0000230 2020 2020 2020 2020 2020 2020 2020 2020  
*  
0000290 2020 2020 2020 2020 0042 0002 04b0 0161  
00002a0 0002 0000 013d 0002 0001 009c 0002 000e  
00002b0 0019 0002 0000 0012 0002 0000 0013 0002  
00002c0 0000 01af 0002 0000 01bc 0002 0000 003d  
00002d0 0012 0000 000f 3f1b 27f6 0038 0000 0000  
00002e0 0001 0258 0040 0002 0000 008d 0002 0000  
00002f0 0022 0002 0000 000e 0002 0001 01b7 0002
```

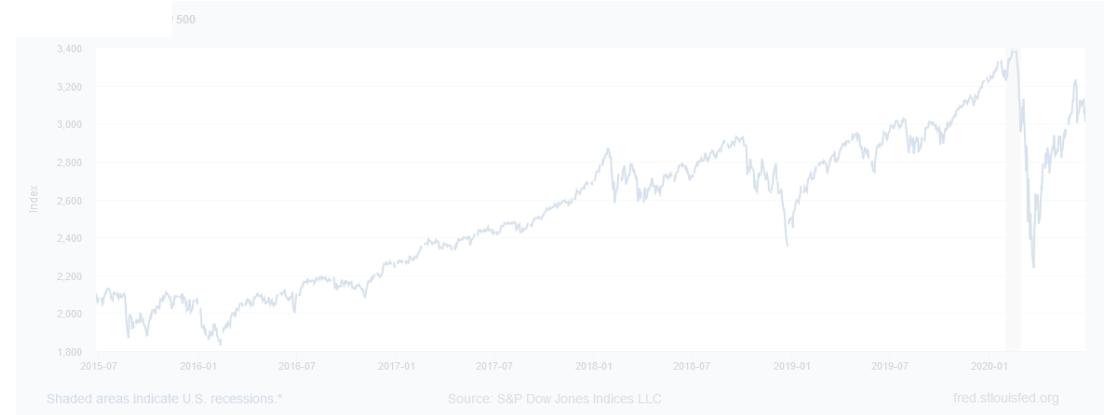


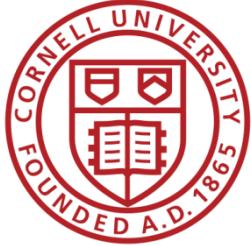
| |
|---------|
| 2101.49 |
| 2057.64 |
| 2063.11 |
| 2077.42 |
| 2076.78 |
| 0 |
| 2068.76 |
| 2081.34 |
| 2046.68 |
| 2051.31 |
| 2076.62 |
| 2099.60 |
| 2108.95 |
| 2107.40 |
| 2124.29 |
| 2126.64 |
| 2128.28 |
| 2119.21 |
| 2114.15 |
| 2102.15 |
| 2079.65 |
| 2067.64 |
| 2093.25 |
| 2108.57 |
| 2108.63 |
| 2103.84 |



“It’s a file called SP500.xlsx”

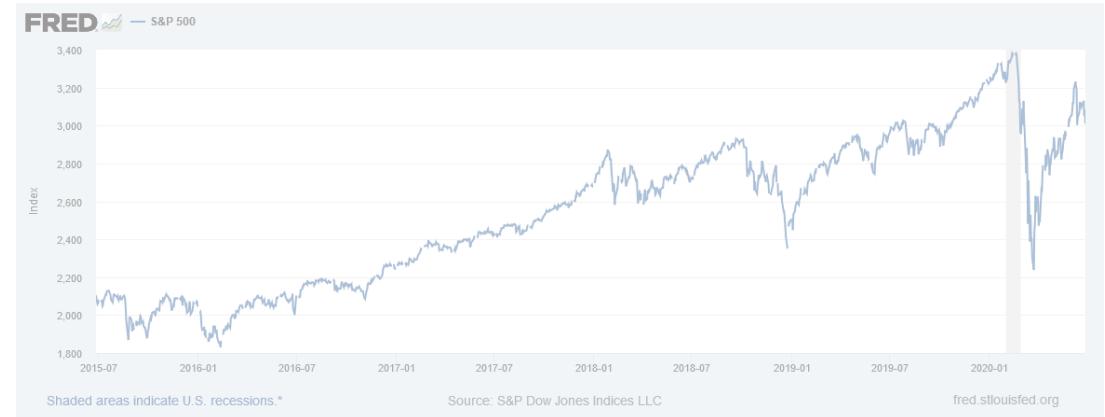
| SP500 | S&P 500, Index, Daily, Not Seasonally Adjusted |
|-------------------------|---|
| Frequency: Daily, Close | |
| observation_date | SP500 |
| 2015-06-26 | 2101.49 |
| 2015-06-29 | 2057.64 |
| 2015-06-30 | 2063.11 |
| 2015-07-01 | 2077.42 |
| 2015-07-02 | 2076.78 |
| 2015-07-03 | 0 |
| 2015-07-06 | 2068.76 |
| 2015-07-07 | 2081.34 |
| 2015-07-08 | 2046.68 |
| 2015-07-09 | 2051.31 |
| 2015-07-10 | 2076.62 |
| 2015-07-13 | 2099.60 |
| 2015-07-14 | 2108.95 |
| 2015-07-15 | 2107.40 |
| 2015-07-16 | 2124.29 |
| 2015-07-17 | 2126.64 |
| 2015-07-20 | 2128.28 |

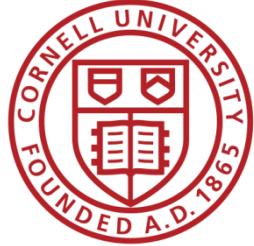




“It’s a file called SP500.xlsx, downloaded from FRED.”

| SP500 | S&P 500, Index, Daily, Not Seasonally Adjusted |
|-------------------------|---|
| Frequency: Daily, Close | |
| observation_date | SP500 |
| 2015-06-26 | 2101.49 |
| 2015-06-29 | 2057.64 |
| 2015-06-30 | 2063.11 |
| 2015-07-01 | 2077.42 |
| 2015-07-02 | 2076.78 |
| 2015-07-03 | 0 |
| 2015-07-06 | 2068.76 |
| 2015-07-07 | 2081.34 |
| 2015-07-08 | 2046.68 |
| 2015-07-09 | 2051.31 |
| 2015-07-10 | 2076.62 |
| 2015-07-13 | 2099.60 |
| 2015-07-14 | 2108.95 |
| 2015-07-15 | 2107.40 |
| 2015-07-16 | 2124.29 |
| 2015-07-17 | 2126.64 |
| 2015-07-20 | 2128.28 |



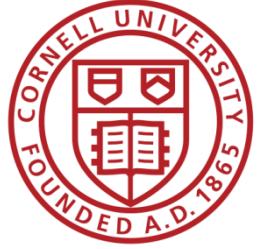


“It’s a file called SP500.xlsx, downloaded from FRED.”

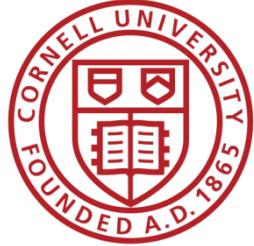
| | |
|---|---|
| SP500 | S&P 500, Index, Daily, Not Seasonally Adjusted |
| Frequency: Daily, Close observation_date | SP500 |
| 2015-06-26 | 2101.49 |
| 2015-06-29 | 2057.64 |
| 2015-06-30 | 2063.11 |
| 2015-07-01 | 2077.42 |
| 2015-07-02 | 2076.78 |
| 2015-07-03 | 0 |
| 2015-07-06 | 2068.76 |
| 2015-07-07 | 2081.34 |
| 2015-07-08 | 2046.68 |
| 2015-07-09 | 2051.31 |
| 2015-07-10 | 2076.62 |
| 2015-07-13 | 2099.60 |
| 2015-07-14 | 2108.95 |
| 2015-07-15 | 2107.40 |
| 2015-07-16 | 2124.29 |
| 2015-07-17 | 2126.64 |
| 2015-07-20 | 2128.28 |

S&P Dow Jones Indices LLC. 2020. “*S&P 500 [SP500] [dataset]*”, retrieved from FRED, Federal Reserve Bank of St. Louis; <https://fred.stlouisfed.org/series/SP500>, June 26, 2020.





We're all set then.

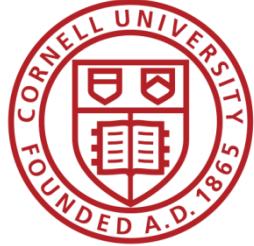


“SP500.xlsx, from S&P (2020). Not provided as part of replication package because © S&P.”

| SP500 | S&P 500, Index, Daily, Not Seasonally Adjusted |
|-------------------------|---|
| Frequency: Daily, Close | |
| observation_date | SP500 |
| 2015-06-26 | 2101.49 |
| 2015-06-29 | 2057.64 |
| 2015-06-30 | 2063.11 |
| 2015-07-01 | 2077.42 |
| 2015-07-02 | 2076.78 |
| 2015-07-03 | 0 |
| 2015-07-06 | 2068.76 |
| 2015-07-07 | 2081.34 |
| 2015-07-08 | 2046.68 |
| 2015-07-09 | 2051.31 |
| 2015-07-10 | 2076.62 |
| 2015-07-13 | 2099.60 |
| 2015-07-14 | 2108.95 |
| 2015-07-15 | 2107.40 |
| 2015-07-16 | 2124.29 |
| 2015-07-17 | 2126.64 |
| 2015-07-20 | 2128.28 |

S&P Dow Jones Indices LLC. 2020. “*S&P 500 [SP500] [dataset]*”, retrieved from FRED, Federal Reserve Bank of St. Louis; <https://fred.stlouisfed.org/series/SP500>, June 26, 2020.





Data Availability Statements

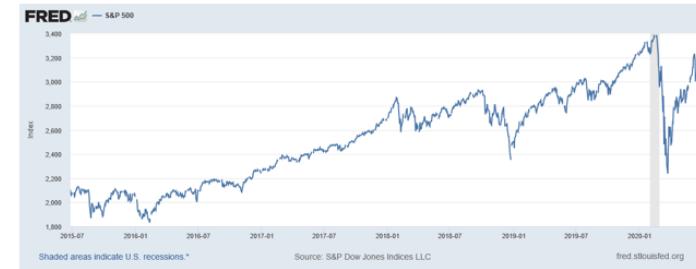
Describes data file, where to get it, how to get it, and any conditions of obtaining it

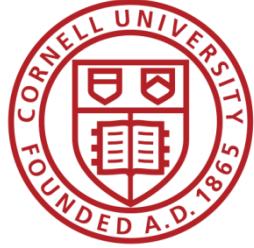
| | |
|------------|---------|
| 2015-07-15 | 2107.40 |
| 2015-07-16 | 2124.29 |
| 2015-07-17 | 2126.64 |
| 2015-07-20 | 2128.28 |

“SP500.xlsx, from S&P (2020). Not provided as part of replication package because © S&P.”

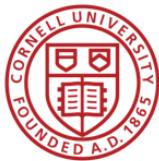
S&P 500
S&P 500, Index, Daily,
Not Seasonally Adjusted

S&P Dow Jones Indices LLC. 2020. “S&P 500 [SP500] [dataset]”, retrieved from FRED, Federal Reserve Bank of St. Louis; <https://fred.stlouisfed.org/series/SP500>, June 26, 2020.





Data Citation



“SP500.xlsx, from S&P (2020). Not provided as part of replication package because © S&P.”

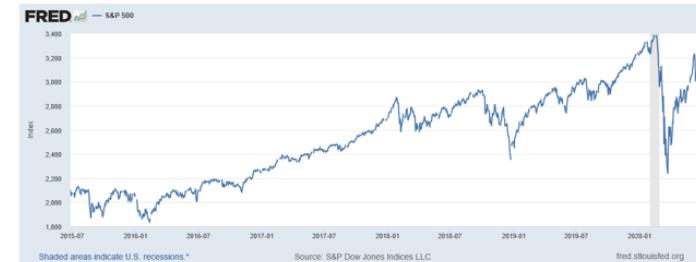
Attributes the file to
the proper source

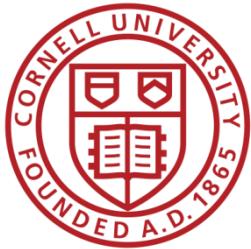
SP500

S&P 500, Index, Daily,
Not Seasonally
adjusted

| Date | Value |
|------------|---------|
| 2015-07-08 | 2101.49 |
| 2015-07-09 | 2057.64 |
| 2015-07-10 | 2063.11 |
| 2015-07-13 | 2074.42 |
| 2015-07-14 | 2076.78 |
| 2015-07-15 | 0 |
| 2015-07-16 | 2068.76 |
| 2015-07-17 | 2081.34 |
| 2015-07-20 | 2046.68 |

S&P Dow Jones Indices LLC. 2020. “S&P 500 [SP500] [dataset]”, retrieved from FRED, Federal Reserve Bank of St. Louis; <https://fred.stlouisfed.org/series/SP500>, June 26, 2020.





perceived criteria of importance.

1. Importance

Data should be considered legitimate, citable products of research. Data should be accorded the same importance in the scholarly record as citat research objects, such as publications[1].



Data Citation Principles

2. Credit and Attribution

Data citations should facilitate giving scholarly credit and normative and le attribution to all contributors to the data, recognizing that a single style or of attribution may not be applicable to all data[2].

3. Evidence

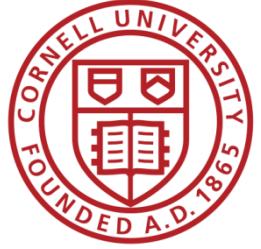
In scholarly literature, whenever and wherever a claim relies upon data, the corresponding data should be cited[3].

4. Unique Identification

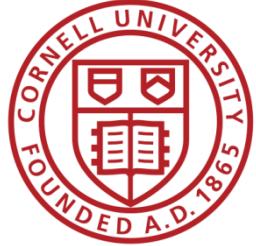
A data citation should include a persistent method for identification that i actionable, globally unique, and widely used by a community[4].

5. Access

Data citations should facilitate access to the data themselves and to such metadata, documentation, code, and other materials as are necessary for

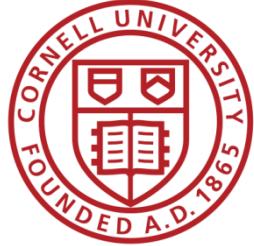


Data Citations are Hard!



At the AEA, every manuscript is checked

- What datasets are used
- Are they cited?
- Is there additional information on access?
- Is there license/ data use information?
 - → Should the author provide the data?
 - → Is the author allowed to provide data?



Example 2: Academic data publisher

 **ECONOMIC POLICY UNCERTAINTY**

Home Methodology Media Research & Applications About Us

EPU Indices

[All Country-Level Data](#)

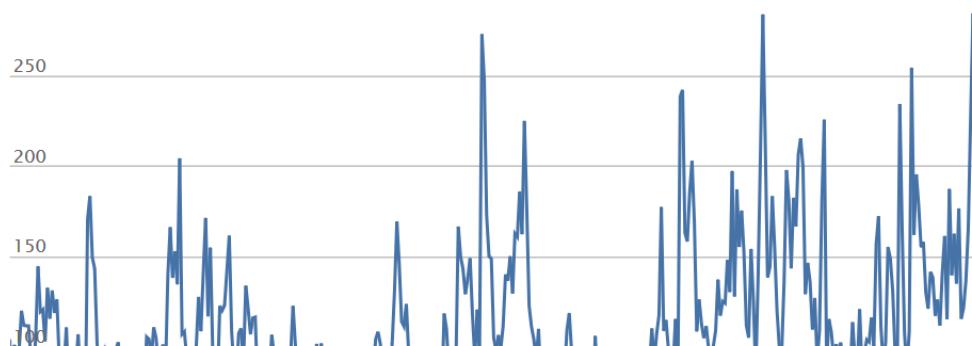
| | |
|-----------------------------|-----------------------------|
| Global | USA |
| Australia | Brazil |
| Canada | Chile |
| China | Colombia |
| Croatia New | France |
| Germany | Greece |
| Hong Kong | India |
| Ireland | Italy |
| Japan | South Korea |

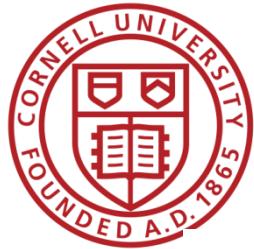
Economic Policy Uncertainty Index

We develop indices of economic policy uncertainty for countries around the world.

Monthly US Economic Policy Uncertainty Index

Zoom [1m](#) [3m](#) [6m](#) [1y](#) [7y](#) [All](#)





Example 2: Academic data publisher

https://www.policyuncertainty.com/index.html SEP DEC JAN
103 captures 14 2018 2019 2020
18 Aug 2012 - 14 Dec 2019

ECONOMIC POLICY UNCERTAINTY [Home](#) [Methodology](#) [Media](#) [Research & Applications](#) [About Us](#)

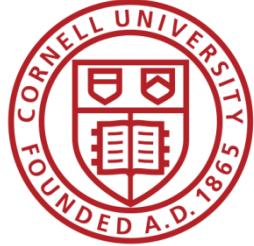
EPU Indices **Economic Policy Uncertainty Index**

All Country-Level Data We develop indices of economic policy uncertainty for countries around the world.

Globe Australia Canada China Croatia France Germany Greece Hong Kong India Ireland Italy Japan South Korea

© 2012-2018 by Economic Policy Uncertainty

The chart displays a highly volatile line graph representing the Economic Policy Uncertainty Index. The y-axis ranges from 100 to 200, and the x-axis spans from approximately 2012 to 2018. The index fluctuates significantly, with major peaks around 2013, 2015, and 2017, and troughs around 2014, 2016, and 2018.



Example 2: Academic data publisher-new!

 **ECONOMIC POLICY UNCERTAINTY**

[Home](#) [Methodology](#) [Media](#) [Research & Applications](#) [About Us](#)

[EPU Indices](#)

All Country-Level Data

Global [USA](#)

[Monthly US Economic Policy Uncertainty Index](#)

We develop indices of economic policy uncertainty for countries around the world.

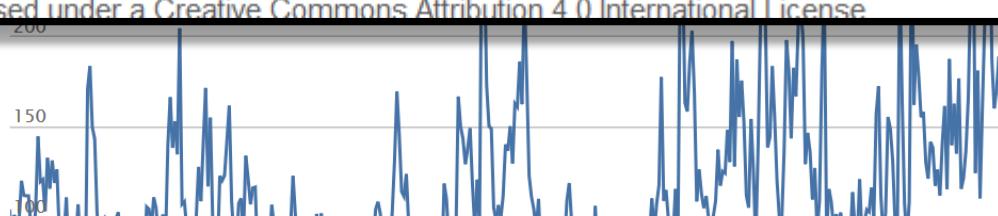
This work is licensed under a Creative Commons Attribution 4.0 International License

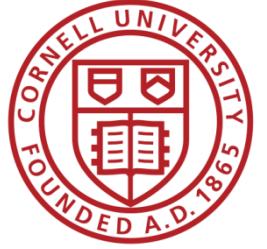
Germany [Greece](#)

[Hong Kong](#) [India](#)

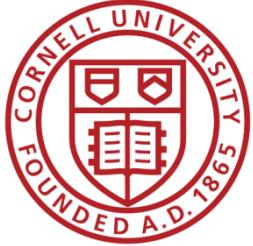
[Ireland](#) [Italy](#)

[Japan](#) [South Korea](#)



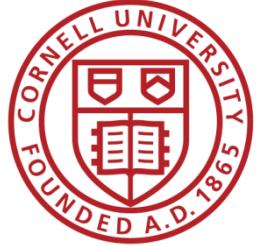


Data Rights



Rights to use data

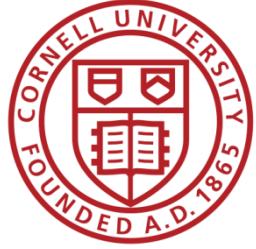
- You browsed a website
- You purchased the data
- You signed a data use agreement
- You created the data (lab experiment)
- You had survey respondents consent to use (IRB approval!)



Rights to distribute the data

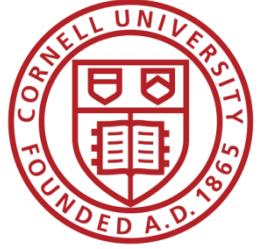
- If you created the data, you decide.
- If you got it from somewhere else:

READ THE TERMS OF USE / DATA USE
AGREEMENT / CLICK-THROUGH / ETC.

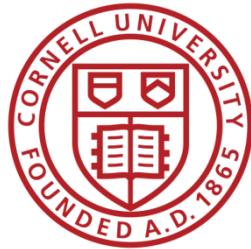


Citing restricted-access data

“Well, I can’t download the data, so I can’t cite it.”



Wrong



Example 4: German Restricted-access



RESEARCH DATA CENTRE (FDZ)
of the German Federal Employment Agency (BA)
at the Institute for Employment Research (IAB)

[Home](#) | [Newsletter](#) | [Jobs](#) | [Contact](#) | [Data Privacy](#) | [Imprint](#)



| Data Version | DOI (Link to Description of Data Version) | Availability (yyyy-mm-dd) |
|------------------------------|---|---------------------------|
| BHP 7518 v1 (current) | 10.5164/IAB.BHP7518.de.en.v1 | 2020-01-13 |
| BHP 7517 v1 | 10.5164/IAB.BHP7517.de.en.v1 | 2018-12-12 |
| BHP 7516 v1 | 10.5164/IAB.BHP7516.de.en.v1 | 2018-04-11 |

External data

Data Archive

Data Access

Campus Files

Publications

Events

Projects of FDZ users

FDZ Projects

Complaint point of the
RatSWD

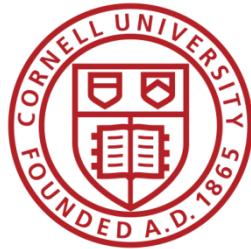
Figures of the FDZ

employees, both in total and broken down by gender, age, occupational status, qualification and nationality. Means and medians of wages for full-time employees are given, too. Additional datasets providing information about (gross) worker flows and about foundations and closures of establishments are available on request.

Data Versions

Old versions are only available for replication studies and only in justified exceptional cases for new Projects.

| Data Version | DOI (Link to Description of Data Version) | Availability (yyyy-mm-dd) |
|------------------------------|---|---------------------------|
| BHP 7518 v1 (current) | 10.5164/IAB.BHP7518.de.en.v1 | 2020-01-13 |



Example 4: German Restricted-access



RESEARCH DATA CENTRE (FDZ)
of the German Federal Employment Agency (BA)
at the Institute for Employment Research (IAB)

[Home](#) | [Newsletter](#) | [Jobs](#) | [Contact](#) | [Data Privacy](#) | [Imprint](#)



| Data Version | DOI (Link to Description of Data Version) | Availability (yyyy-mm-dd) |
|------------------------------|---|---------------------------|
| BHP 7518 v1 (current) | 10.5164/IAB.BHP7518.de.en.v1 | 2020-01-13 |
| BHP 7517 v1 | 10.5164/IAB.BHP7517.de.en.v1 | 2018-12-12 |
| BHP 7516 v1 | 10.5164/IAB.BHP7516.de.en.v1 | 2018-04-11 |

External data

Data Archive

Data Access

Campus Files

Publications

Events

Projects of FDZ users

FDZ Projects

Complaint point of the
RatSWD

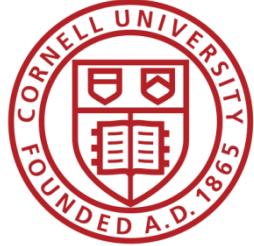
Figures of the FDZ

employees, both in total and broken down by gender, age, occupational status, qualification and nationality. Means and medians of wages for full-time employees are given, too. Additional datasets providing information about (gross) worker flows and about foundations and closures of establishments are available on request.

Data Versions

Old versions are only available for replication studies and only in justified exceptional cases for new Projects.

| Data Version | DOI (Link to Description of Data Version) | Availability (yyyy-mm-dd) |
|------------------------------|---|---------------------------|
| BHP 7518 v1 (current) | 10.5164/IAB.BHP7518.de.en.v1 | 2020-01-13 |



Example 4: German Restricted-access

Establishment History Panel (BHP) – Version 7518 v1

DOI: 10.5164/IAB.BHP7518.de.en.v1

Summary

Data source:

Data Access

The IAB Establishment Panel is available via the following ways of access:

- On-site use at the FDZ. Further information on Applying for [on-site use](#).
- Remote data Access. Further information on Applying for [remote data access](#).

nationality. Means and medians of wages for full-time employees are given, too. Additional datasets providing information about (gross) worker flows and about foundations and closures of establishments are available on request.

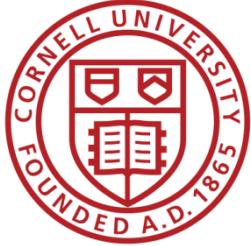
Dataset Descriptions and Frequencies

German

- DOI: [10.5164/IAB.FDZD.2001.de.v1](https://doi.org/10.5164/IAB.FDZD.2001.de.v1)
- [FDZ-Datenreport 01/2020](#)
- [Fallzahlen und Labels](#)

English

- DOI: [10.5164/IAB.FDZD.2001.en.v1](https://doi.org/10.5164/IAB.FDZD.2001.en.v1)



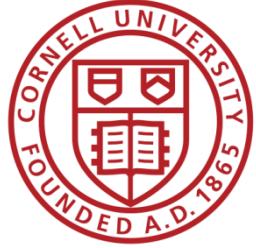
Element of a (data) citation

- CMOS notes (16th ed., 14.1)

“Regardless of the convention being followed, **source citations** must always provide sufficient information either to **lead readers directly to the sources** consulted or, for materials that may ***not be readily available***, to enable readers to ***positively identify them, regardless of whether the sources are published or unpublished or in printed or electronic form.***”

ICPSR notes that a citation should include the following items:

- Author
- Title
- Distributor
- Date
- Version
- Persistent identifier



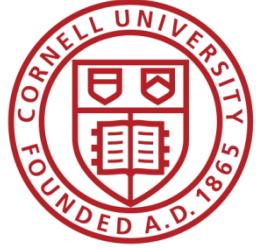
These are archives



Source: <https://chiddicksfamilytree.com/2017/11/08/use-it-or-lose-it/>

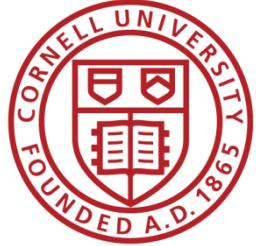


Source: <https://www.archives.gov/publications/prologue/2008/spring/frc.html>



They are citable

- Typescript of short story Brothers and Sisters by Budge Wilson, 2000, MS-2-650.2013-070, Box 3, Folder 9, Budge Wilson fonds, Dalhousie University Archives, Halifax, Nova Scotia, Canada.



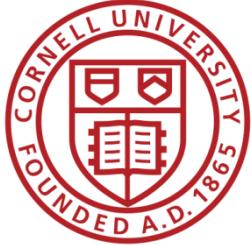
Element of a (data) citation

ICPSR notes that a citation should include the following items:

- Author
- Title
- Distributor
- Date
- Version
- Persistent identifier

Suggested Citation:

S&P Dow Jones Indices LLC, *S&P 500 [SP500]*, retrieved from FRED, Federal Reserve Bank of St. Louis;
<https://fred.stlouisfed.org/series/SP500>, June 26, 2020.



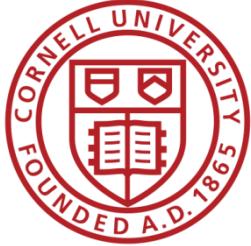
Element of a (data) citation

ICPSR notes that a citation should include the following items:

- Author
- Title
- Distributor
- Date
- Version
- Persistent identifier

Constructed Citation:

Institute for Employment Research (IAB), Establishment History Panel 1975-2018. Accessed via the Research Data Centre (FDZ) of the German Federal Employment Agency DOI: 10.5164/IAB.BHP7518.de.en. v1 June 26, 2020.

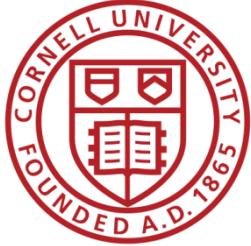


Element of a (data) citation

ICPSR notes that a citation should include the following items:

- Author
- Title
- Distributor
- Date
- Version
- Persistent identifier

Constructed Citation:
US Census Bureau,
Longitudinal Business
Database (LBD) 1975-
2018. Last accessed via
the Federal Statistical
Research Data Centre
(FSRDC) June 26, 2020.



And we check them!

- If the URL does not work, we make a note.
- If the site requires registration, we try it out.
 - How long?
 - Any requirements?

- What does the site say?

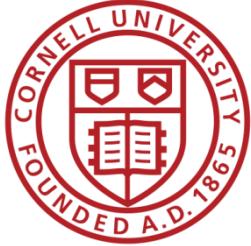
Please use the following citation when referring to this file in the different versions:

Inglehart, R., C. Haerpfer, A. Moreno, C. Welzel, K. Kizilova, J. Diez-Medrano, M. Lagos, P. Norris, E. Ponarin & B. Puranen et al. (eds.). 2014. World Values Survey: Round Six - Country-Pooled Datafile Version:

www.worldvaluessurvey.org/WVSDocumentationWV6.jsp.

Madrid: JD Systems Institute.

- Is that in the README / Paper/ Appendix?



And we

In order to download the file you are as
on the "Conditions of Use". Please read

PERSONAL DATA

Title (position):

Full name:

Company/Institution:

E-mail:

FILE USAGE

Project title:

Intended use:

Brief description of the purpose of applic

CONDITIONS OF USE

1. Restrictions

These data files are available without restrictions, provided

- a) that they are used for non-profit purposes; and
- b) correct citations are provided and sent to the World Values Survey Association for each publication of results based in part or entirely on these data files. This citation will be made freely available; and
- c) the data files themselves are not redistributed.

2. Correct citation

CONDITIONS OF USE

1. Restrictions

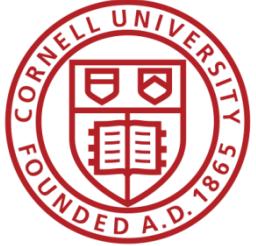
These data files are available without restrictions, provided

- a) that they are used for non-profit purposes; and
- b) correct citations are provided and sent to the World Values Survey Association for each publication of results based in part or entirely on these data files. This citation will be made freely available; and
- c) the data files themselves are not redistributed.

2. Correct citation

Madrid. JD Systems Institute.

- Is that in the README / Paper/ Appendix?
- Are all the conditions met/described?



Data Availability

- A statement about **data availability**
 - DOI assigned
 - But longer
- A statement about **usage rights**
 - Not every dataset is in the public domain
 - Not everybody knows that U.S. Government data are usually in the public domain



Data Availability Statements (DAS)

- A statement about **where data** supporting the results reported in a published article can be

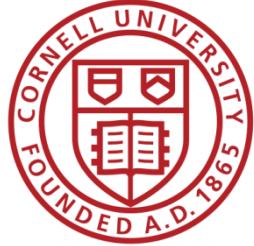
o publicly
ated during

y providing a

I restrictions,

Provide data citations (in manuscript) and data availability statements (in README or appendix)

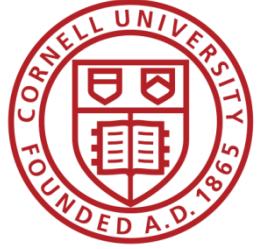
Take-away



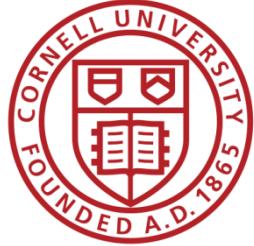
Data: Citations, Access, Rights

- Any data can be cited – even if you can't download it
- Any data that you accessed ... can have that access be described
 - But caution: It should be such that others can also repeat the access!
- Just because you “have” the data does not mean you can give it to others
 - Also: distinguish between “sharing” and “publishing”
 - Know your terms of use!

Documenting



What about your code?



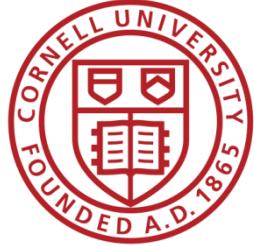
How is your code documented?

For yourself:

- What are you doing on lines 30-45?
- Will you still remember that in 3 years?
- Does it require any instructions outside of the code?

For your audience:

- Your thesis advisor
- Your boss
- Your co-worker, to whom you've been trying to hand off this project for two weeks and you've finally convinced your boss...



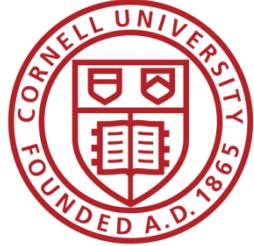
Poor coding practices

- **Manual/non-automation**

Code produces no meaningful output

- **Lack of robustness:**

Bugs in the code

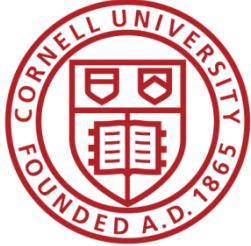


Basic coding hints

Listing 1: mystuff.sas

```
1  data  "C:\ Users\Me\CensusChina.sas7bdat";
2      set  "C:\ Users\Me\CensusChina.sas7bdat";
3      earn=log(earn);
4  run;
5  proc reg data="C:\ Users\Me\CensusChina.sas7bdat";
6  model earn = sex education experience;
7  run;
```

What can possibly be wrong about that?



What's wrong with that?

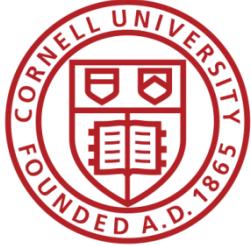
Everything!

- *Name of program*: uninformative
- *Destruction of original data*:
program cannot be re-run for same
results
- *No portability*: cannot be run
anywhere else
- *No explanation*: why are we doing
this?

Listing 1: mystuff.sas

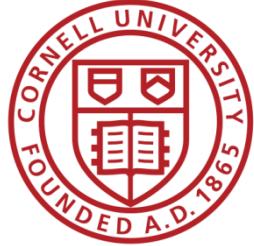
```
1 data "C:\Users\Me\CensusChina.sas7bdat";
2 set "C:\Users\Me\CensusChina.sas7bdat";
3 earn=log(earn);
4 run;
5 proc reg data="C:\Users\Me\CensusChina.sas7bdat";
6 model earn = sex education experience;
7 run;
```

What can possibly be wrong about that?



Structuring programs

- **Portability:** should be easy to adjust to a new location
 - Your new laptop
 - A replicator 10 years down the road
- **Clarity:** don't assume that you will remember what seems so obvious now
 - Describe **what** you are doing, and **why**



File structure

mystuff.R

read.R

version2.R

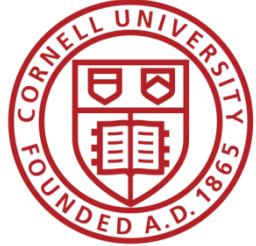
ols.sas

readCensus.R

readBLS.R

prepareCensus.R

runols.sas



Better

01_01_readBLS.R

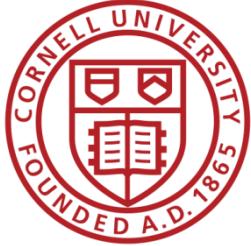
02_01_readCensus.R

02_02_prepareCensus.R

03_01_createanalysisdata.R

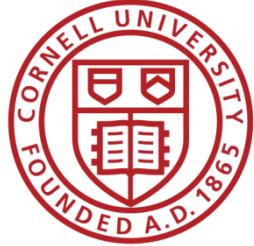
04_01_runOLS.sas

README.txt



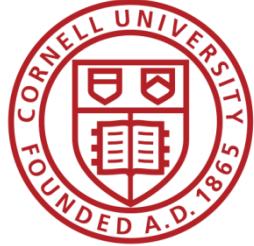
Functions

- Create re-usable code
 - ... for other projects
 - ... within the same project (robustness checks, variation on regressions, etc.)
- Feasible in any programming language
 - Stata: “program” or ado files
 - SAS: macros
 - R: functions or packages



Streamlining replication packages

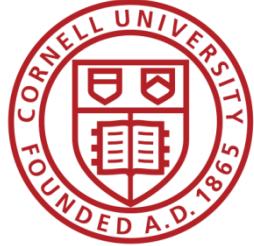
- Master script preferred
 - Least amount of manual effort
- No manual manipulation
 - “Change the parameter to 0.2,
then run the code again” 
- No manual copying of results
 - Write out/save tables and figures
using packages
 - Compute all numbers in package
- No manual install of packages
 - Use a script to create all
directories, install all necessary
packages/requirements/etc.
- Clear instructions!



Two magic Stata lines

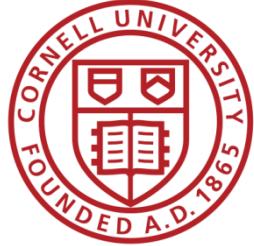
```
global basepath "/my/path/here"  
adopath + "$basepath/ado"
```

In addition to “**graph export**”, “**outtab**”, etc.



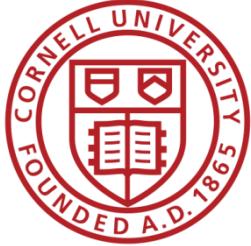
Extreme examples

- Matlab-based simulation
- Real example, 10 figures, 4 panels each
- For Figure 5a, comment line 52, uncomment line 151, run the code, then copy the figure into your document.
- For Figure 5b, comment line 151 again, leave line 52 commented, and change the parameter on line 75 to “3”
-



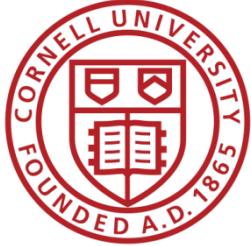
Extreme examples

- Stata-based estimation
- 4 variants
- Run the data creation programs, then copy the data to Folder A
- Copy programs “b.do” and “c.do” from Folder A to Folder B, but modify “c.do” on line 20
- Once done, convert the output from “d.do” to a Matlab file, and run the simulation in Folder B/C
-



Ideal setup

- 1 program to prepare the setup
 - Installs all packages
 - Creates all directories
 - 1 program (or a very small number) that creates the rest
 - Possibly with macros/ ado files/ subroutines
 - Possibly with parameter files that might differ per directory
 - All tables and figures are output programmatically
-
- Setting up can be done in all languages
 - Matlab, Stata, R, Python, Fortran
 - Subroutines exist in all languages
 - You might need to learn how!
 - Ability to output figures and tables (Excel, LaTeX) exist in all languages



How to prepare the replication package

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

AEA Data and Code Guidance



AMERICAN
ECONOMIC
ASSOCIATION

Guidance for authors,
data and code sup-
replicators.

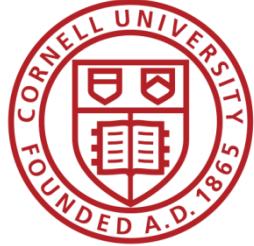
Steps for the Third-party Replicator

- 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 original environment could reasonably rely on the instructions in the

That's our Protocol!

- 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)

Archiving



Full-featured repository

OPEN ICPSR Find Data Share Data openICPSR Repositories ▾

AEA Deposit Instructions **Browse AEA Deposits** **Contact**

Depositing Data in the AEA Data and Code Repository

The *American Economic Association journals* require authors to deposit data and materials with a community-recognized or general repositories. The *AEA Data and Code Repository at ICPSR* serves that purpose. Please see the AEA's [Data and Code Availability Policy](#) and data citation guidance at the [Sample References](#) page for more details. **Authors are required to include a citation pointing to the deposit in the reference section of the final version of the article sent to the AEA.** The *openICPSR* repository automatically generates a citation when the data are "published."

Deposits should include all data, annotated program code, command files, and documentation that is needed to replicate the findings from the authors' submitted article.

- **Data** should be comprehensively documented (see ICPSR's [Guide to Social Science Data Preparation and Archiving, 5th Edition](#) for guidance). The **author** is responsible for removing identifying information from the data to protect [confidentiality](#). Neither the AEA nor ICPSR review submissions for disclosure risk.
- **Program** code and command files should be annotated to facilitate replication and ensure clear correspondence between code and figures, tables, and analyses in the published article.
- Authors retain ownership and copyright to the data and code. Authors are required to affirm that they have the right to publish and redistribute the material. However,
 - ICPSR requires a license for distribution of data.
 - An **open license** is required by the AEA, in order to allow others to re-use the data and code, in particular for replication. Authors can select from several license options, including CC-BY 4.0 for data and Modified BSD for software and code. If an author would like to use multiple licenses or create a customized license, she should select the "Other" license option and upload a LICENSE file alongside the data and documentation.

By depositing in the AEA Data and Code Repository, the depositors allow the AEA staff to add keywords and other metadata which are important for proper indexing in linking. Any other changes are subject to the license chosen for the materials.

[View more extensive \(unofficial\) guidance.](#)

Start Your Deposit



Dépôt institutionnel

 zendesk

Search

UQAM > Archipel

OPENICPSR

[Find Data](#) [Share Data](#) [Repositories](#)

Biblioth

Find and share social, behavioral, and health sciences research data.

immigration

Q

Re

MON

Out

Nouw



AMERICAN
ECONOMIC
ASSOCIATION

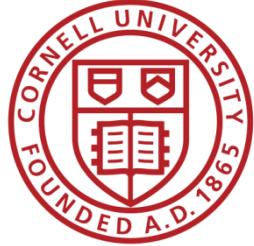


JEH



es coll.

Centre de recherche en éthique [419]



Action: Data citations and metadata

What is FAIR?

- Findable,
- Accessible,
- Interoperable, and
- Re-usable

The FORCE11 logo features a blue circular icon with a white target-like pattern next to the word "FORCE11". Below it is the tagline "The Future of Research Communications and e-Scholarship". A navigation bar below the logo includes "ABOUT", "COMMUNITY", and "CODE OF CON".

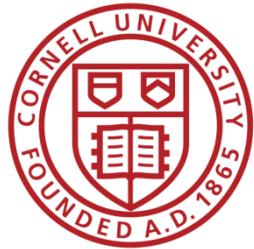
FORCE11 » Groups » The FAIR Data Principles

THE FAIR DATA PRINCIPLES

JOIN IN THE DISCUSSION - LEARN
FAIR Data Principles

Preamble

One of the grand challenges of data-intensiv



FAIR data principles rely on metadata

— Scope of Project

Subject Terms ?

Do not copy/paste multiple terms into this field. Terms must be entered individually.

[✖ Russia](#) [✖ Industry](#) [✖ Factories](#) [✖ Russian Empire](#) [✖ Corporations](#)

JEL Classification ?

[✖ L20 General](#) [✖ N63 Europe: Pre-1913](#) [✖ O43 Institutions and Growth](#)

Manuscript Number ?

AER-2015-1656.R3 [edit](#) [remove](#)

Geographic Coverage ? [+ add value](#)

European Russia (Russian Empire) [edit](#) [remove](#)

Time Period(s) ? [+ add value](#)

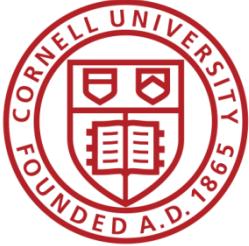
1894 – 1908 (Three years: 1894, 1900, and 1908) [edit](#) [remove](#)

Collection Date(s) ? [+ add value](#)

Universe ?

Manufacturing establishments in the European part of the Russian Empire. [edit](#) [remove](#)

Data Type(s) ?

[Find Data](#) / [Imperial Russian Factory Database, 1894-1908](#)

Imperial Russian Factory Database, 1894-1908

Principal Investigator(s): Amanda Gregg, Middlebury College

Version: V1



AMERICAN
ECONOMIC
ASSOCIATION

| Name | File Type | Last Modified |
|------------------------------------|---|----------------------------|
| 1894MicroData.xlsx | application/vnd.openxmlformats-officedocument.spreadsheetml.sheet | 4.5 MB 08/08/2019 11:01:AM |

Project Citation:

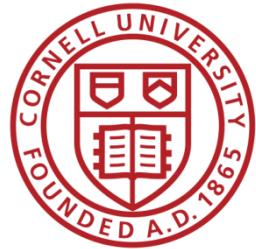
Gregg, Amanda. Imperial Russian Factory Database, 1894-1908. Nashville, TN: American Economic Association [publisher], 2020. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2020-01-29. <https://doi.org/10.3886/E110681V1>

| | | | |
|--|---------------------|---------|---------------------|
| AG_Corp_CleaningandDatabaseCompiler.do | text/x-stata-syntax | 23.4 KB | 08/08/2019 11:02:AM |
|--|---------------------|---------|---------------------|

Related Publications

The following publications are supplemented by the data in this project.

- Gregg, Amanda. "Factory Productivity and the Concession System of Incorporation in Late Imperial Russia, 1894-1908." *American Economic Review* 110, no. 2 (February 2020): 401-27. <https://doi.org/10.1257/aer.20151656>.

[Find Data](#) / [Imperial Russian Factory Database, 1894-1908](#)

Imperial Russian Factory Database, 1894-1908

Principal Investigator(s): Amanda Gregg, Middlebury College

Version: V1

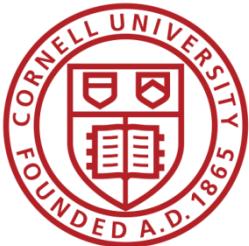


```
<meta name="DC.identifier" content="10.3886/E110681V1" />
<meta name="DC.title" content="Imperial Russian Factory Database, 1894-1908" />

<meta name="DC.creator" content="Amanda Gregg, Middlebury College" />

<meta name="DC.publisher" content="Inter-university Consortium for Political and Social Research (ICPSR)" />
<meta name="DC.date" content="2020-01-29" />
<meta name="DC.type" content="Dataset" />
```

| | | | MB | |
|--|--|---|---------|---------------------|
| | officedocument.spreadsheetml.sheet | | | 08:53:AM |
| | 1908MicroData.xlsx | application/vnd.openxmlformats-officedocument.spreadsheetml.sheet | 2.3 MB | 08/07/2019 11:06:AM |
| | AG_Corp_CleaningandDatabaseCompiler.do | text/x-stata-syntax | 23.4 KB | 08/08/2019 11:02:AM |
| | AG_Corp_Prod_AppendixCode.do | text/x-stata-syntax | 42.2 KB | 12/09/2019 09:19:AM |
| | AG_Corp_Prod_Code.do | text/x-stata-syntax | 26.6 KB | 12/12/2019 03:01:AM |
| | AG_Corp_Prod_Database.dta | application/x-stata | 11 MB | 08/07/2019 08:55:AM |
| | | application/x-stata | 11.9 | 10/08/2014 |



[Find Data](#) / Imperial Russian Factory Database, 1894-1908

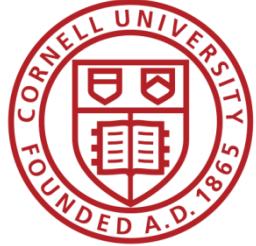
Imperial Russian Factory Database, 1894-1908

Principal Investigator(s):  Amanda Gregg, Middlebury College



```
<script type="application/ld+json">
  {"name": "Imperial Russian Factory Database, 1894-1908", "identifier": "http://doi.org/10.3886/E110681V1", "description": "This database digitizes manufacturing censuses. For each factory, the database includes industry, province, enterprise form, total workers, total revenue, and identifiers that span 908 years also include information on the factory's total machine power. The dataset was constructed to study why some Russian firms chose to become part of a consuming concession system. Note that the final analysis files exclude factories located outside of European Russia and, in the main data files, factories.", "url": "http://doi.org/10.3886/E110681V1", "version": "V1", "keywords": ["Russia", "Industry", "Factories", "Russian Empire", "Corporations"], "spatial": "Russia", "temporalCoverage": ["1894-01-01--1908-12-31 (Three years: 1894, 1900, and 1908)"], "creator": [{"name": "Amanda Gregg", "affiliation": ["Middlebury College"]}, {"name": "openICPSR Self-Deposit Archive", "url": "http://www.openicpsr.org/", "@type": "DataCatalog"}, {"funder": [{"name": "Economic History Association", "url": "http://www.eh.net/", "@type": "Organization"}, {"name": "Yale University Department of Sociology, Anthropology, and Economics", "url": "http://soc.yale.edu/", "@type": "Organization"}, {"name": "Yale Economic Growth Center", "url": "http://yaleegc.yale.edu/", "@type": "Organization"}, {"name": "Yale MacMillan Center", "url": "http://macmillan.yale.edu/", "@type": "Organization"}], "fileFormat": "stata", "contentURL": "https://www.openicpsr.org/openicpsr/project/110681/version/V1/download/terms?path=/openicpsr/110681/fcr:versions/V1/AG_Corp_Prod_Database.dta&type=application/x-stata", "encodingFormat": "application/zip"}, {"fileFormat": "stata", "contentURL": "https://www.openicpsr.org/openicpsr/project/110681/version/V1/download/terms?path=/openicpsr/110681/fcr:versions/V1/AG_Corp_RuscorpMasterFile_Cleaned.dta&type=application/x-stata", "encodingFormat": "application/zip"}, {"fileFormat": "stata", "contentURL": "https://www.openicpsr.org/openicpsr/project/110681/version/V1/AG_Corp_Prod_Database_withAktsiz.dta&type=application/x-stata", "encodingFormat": "application/zip"}], "license": "https://creativecommons.org/licenses/by/4.0", "@context": "http://schema.org", "@type": "Dataset"}}
</script>
```

| | | | | |
|---|--|---------------------|---------|---------------------|
|  | AG_Corp_CleaningandDatabaseCompiler.do | | KB | 11:02:AM |
|  | AG_Corp_Prod_AppendixCode.do | text/x-stata-syntax | 42.2 KB | 12/09/2019 09:19:AM |
|  | AG_Corp_Prod_Code.do | text/x-stata-syntax | 26.6 KB | 12/12/2019 03:01:AM |
|  | AG_Corp_Prod_Database.dta | application/x-stata | 11 MB | 08/07/2019 08:55:AM |
|  | | application/x-stata | 11.9 KB | 10/08/2014 09:45:AM |



... and findability relies on metadata

Google



imperial russian factory



1 dataset found



Imperial Russian Factory
Database, 1894-1908

www.openicpsr.org
search.datacite.org
+1more



Updated Jan 29, 2020



Not seeing a result you expected?
[Learn](#) how you can add new
datasets to our index.



AMERICAN
ECONOMIC
ASSOCIATION

Imperial Russian Factory Database, 1894-1908

[Explore at openICPSR](#)

[Explore at search.datacite.org](#)

[Explore at www.da-ra.de](#)

2 scholarly articles cite this dataset ([View in Google Scholar](#))



Unique identifier

<https://doi.org/10.3886/E110681V1>

Dataset updated Jan 29, 2020

Dataset provided by

[American Economic Association](#)

Authors

Amanda Gregg

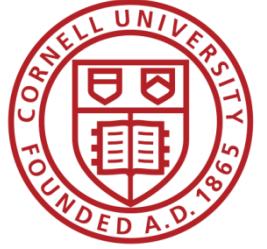
License

[Attribution 4.0 \(CC BY 4.0\)](#)

License information was derived automatically

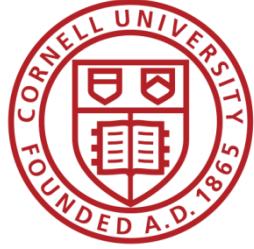
Area covered

European Russia (Russian Empire)



But I've already got it on Github!





Scroll down a bit

Screenshot of a GitHub repository settings page for "labordynamicsinstitute / covid19-expectations-data".

The page includes a header with a GitHub icon, search bar, and navigation links: Pull requests, Issues, Marketplace, Explore, and a user profile icon.

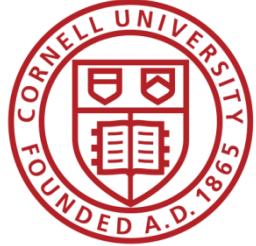
The repository name is "covid19-expectations-data".

The "Settings" tab is selected. On the left, a sidebar shows options: Options (selected), Manage access, Security & analysis, Branches, and a "..." button.

The main content area displays the "Settings" page with the following sections:

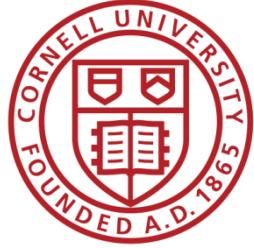
- Repository name:** "covid19-expectations-data" with a "Rename" button.
- Template repository:** An unchecked checkbox with a descriptive link below it.

At the bottom, there is a large red button labeled "Delete this repository". Below it, a note says: "Once you delete a repository, there is no going back. Please be certain."



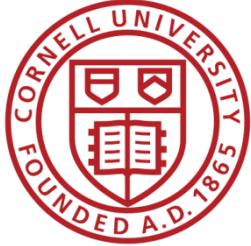
What is not an archive

- Github
- Dropbox
- Your university website
- Your personal website on Google Pages
- ...



And: releases!

The screenshot shows a GitHub repository page for `labordynamicsinstitute / covid19-expectations-data`. The page includes a navigation bar with links to Code, Issues (3), Pull requests, Actions, Projects, Wiki, Security, Insights, and Set. The `Releases` tab is currently selected. Below the tabs, there are fields for `Tag version`, `Target: master`, and a note to choose an existing tag or create a new tag on publish. There is also a `Release title` input field, a `Write` button, a `Preview` button, and a large text area for `Describe this release`.



Suggestions

Use releases as part of your development process

- Define critical moments
 - Showing it to your thesis advisor
 - Submitting to a journal
 - Submitting the R&R
 - Submitting the 5th R&R
 - Responding to the Data Editor
- Hit that “Create release” button
 - Or “`git tag -m “my msg” v20200924`”

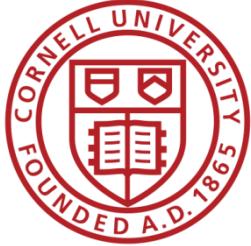
Learn how to archive

- [openICPSR sandbox](#)
- [Zenodo sandbox](#)
- [Github-to-Zenodo API](#)

Archive key (public) moments

- Submitting to a journal (may be required!)
- Presenting at RT2!

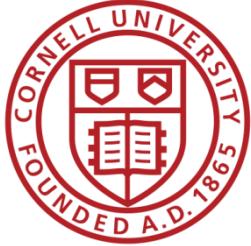
Special Cases



You run your own survey

Welcome, you are a data producer, = Census Bureau!

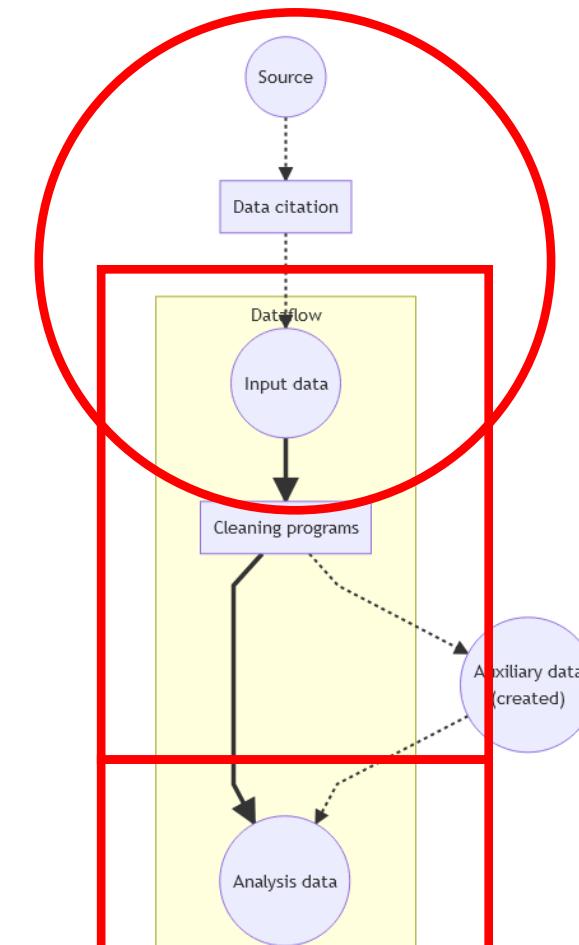
- Your **questionnaires**, your *instructions* (to self, to survey takers), your *sampling scheme*: **metadata**, part of the replication package
- Your **raw survey data**: **untouched**, **archived** (but internally: PII!)
- Your **survey prep code**: stripping identifiers, quasi-identifiers:
archived, (mostly) public

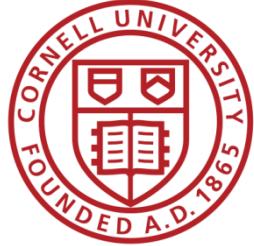


You run your own survey

Welcome, you are a data provider, even for yourself!

- Your (protected, shareable) **survey data**:
 - **Archived**, public (subject to consent)
 - Provided with **license**
 - **Immutable!** Dirty! As-Is!
- You, as anybody else, treats it as “**Source**”
 - Cited!
 - All cleaning happens as part of your replication package!

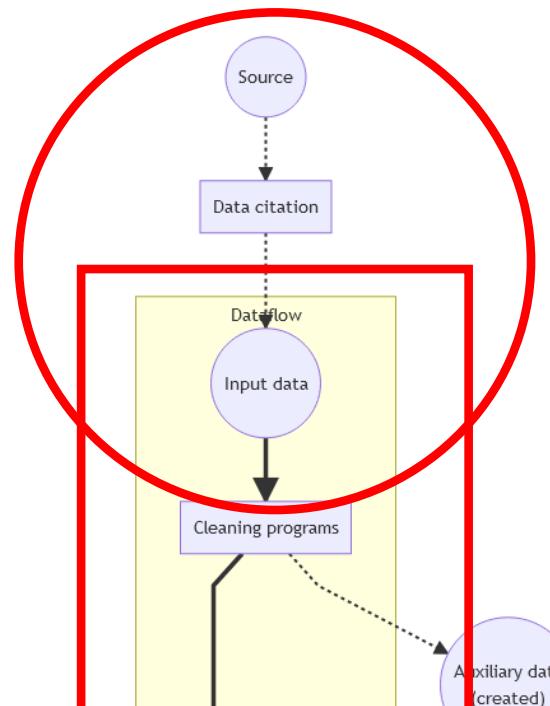


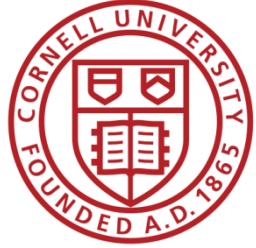


You run an lab experiment

Welcome, you are a data producer, = Census Bureau!

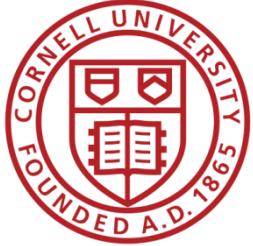
- Etc. etc. etc....
- You provide the “**programming for the experiment**” (zTree, etc.) == survey instrument





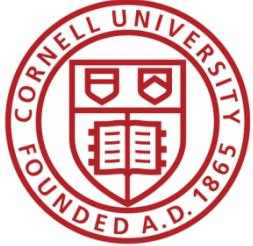
You work in an RDC

- RDC in the sense: restricted-access data environment
- You do not control the environment (storage, backups, etc.)
- You cannot just remove artifacts (data, programs)



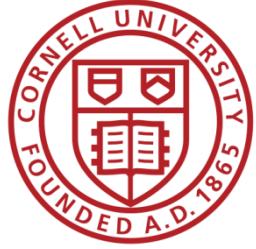
You work in an RDC

- Make your own “github”!
 - Set aside an area of your workspace for “immutable” copies
 - Create “releases” as you request removal of data and code
 - Mirror those “releases” on the outside (at Github, ec.)



You work in an RDC

- Document access process
 - How you got access in the first place
 - How somebody else could get access today
 - That might not be the same!
 - You might not know about it – inform yourself!
 - How somebody can get access to the *files that you cannot remove*



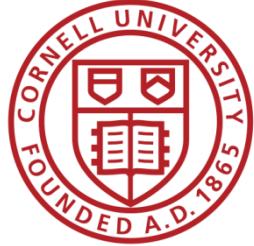
Social science “guild”



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

Thank you!

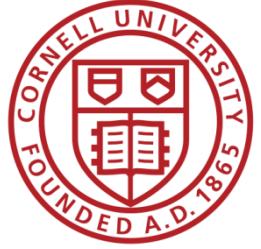
<https://doi.org/10.5281/zenodo.xxxxx>



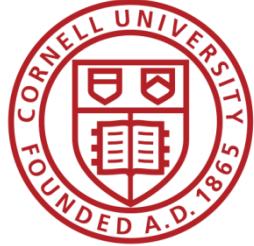
Some resources

- <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)





Addendum – just in...



Robust Science and Open Data

Pinned Tweet

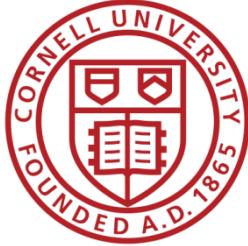
 **Lars Vilhuber** @larsvil · 1h

The @EPA hides behind a fake excuse allowing it to reject scientific findings in an unscientific way. Let me state it clearly: it is possible to do robust, #reproducible, #replicable, open, ethical science with restricted-access data



E.P.A. Rejects Its Own Findings That a Pesticide Harms Children's Brains
The agency's new assessment directly contradicts federal scientists' conclusions five years ago that chlorpyrifos can stunt brain development...
nytimes.com

<https://twitter.com/larsvil/status/1309141822518812676>



What is at issue?

- EPA wants to limit what scientific studies can be taken into consideration when determining rules/policy/etc.

Only Open Access data would count

The screenshot shows a page from the Federal Register. At the top, there are logos for the National Archives and the Federal Register, followed by the title "FEDERAL REGISTER" and the subtitle "The Daily Journal of the United States Government". To the right is the seal of the National Archives and Records Administration. Below the title, a blue bar contains the text "(PR) Prop". The main content area features a large heading "Strengthening Transparency in Regulatory Science". Below it, a sub-headline reads "A Proposed Rule by the Environmental Protection Agency on 03/18/2020". The page is structured with several sections: "PUBLISHED DOCUMENT", "AGENCY:", "ACTION:", "SUMMARY:", and "DOCUMENT DE...". On the far right, there are links for "Printed vers PDF", "Publication 03/18/2020", "Agency: Environmental Protection Agency", and "Dates: Comments or before".

FEDERAL REGISTER
The Daily Journal of the United States Government

(PR) Prop

Strengthening Transparency in Regulatory Science

A Proposed Rule by the Environmental Protection Agency on 03/18/2020

PUBLISHED DOCUMENT

AGENCY:

Environmental Protection Agency (EPA).

ACTION:

Supplemental notice of proposed rulemaking.

SUMMARY:

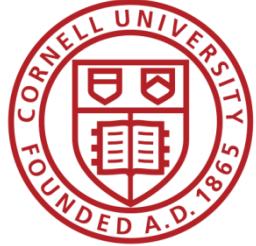
DOCUMENT DE...

Printed vers PDF

Publication 03/18/2020

Agency: Environmental Protection Agency

Dates: Comments or before



But: that's not the criterion for robust science

- A prime example is the government's own **Federal Statistical Research Data Center (FSRDC)** system, where confidential data from **five agencies** (health, taxes, wages, etc.), is used at any time by **several hundred researchers**, on hundreds of approved projects

Federal Statistical Research Data Centers

Some FSRDC locations are temporarily closed.

Please select "Read More" below for additional information.

[Read More](#)



Federal Partners

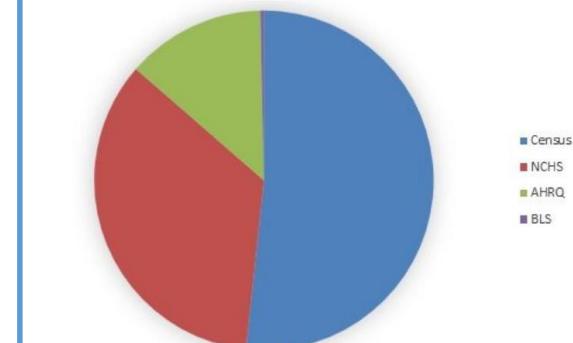


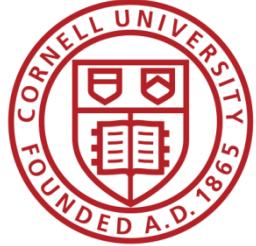
RDC Locations



Active Projects by Federal Partner Agency

As of January 2018, there are 294 approved active research projects in the RDCs. Projects using data provided by the Census Bureau account for 52 percent of total projects. Projects using data provided by the National Center for Health Statistics (NCHS) and the Agency for Healthcare Research and Quality (AHRQ) account for 48 percent of total projects. There is one active BLS project, with several more expected to start in 2018.





True worldwide

- Canada
- France
- UK
- Scandinavia
- Germany
-

≡ C\SD C PROJETS DONNÉES PUBL

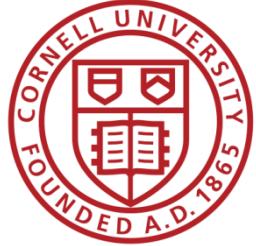
Secure Data Hub

Travail, Emploi
Société, Justice, Éducation
Économie, Entreprises, Finance
Environnement, Agriculture
Santé

A map of Canada color-coded by region: Atlantic (light blue), Prairies (medium blue), Quebec (brown), Ontario (dark brown), and Northwest Territories (grey). Numerous red dots represent 'Centres' and orange dots represent 'Branches' of research data centres across the country.

Find a Research Data Centre

| | |
|----------|-----------------------|
| Atlantic | Prairies |
| Quebec | British Columbia |
| Ontario | Northwest Territories |



**It is possible to do robust, reproducible,
replicable, open, ethical science with
restricted-access data**



<https://twitter.com/larsvil/status/1309141822518812676>