Leaks, Attribution, and Academic Research*

A Case for Professional Reflexivity

Thomas Fox

February 20, 2024

Researchers face serious ethical and professional issues when considering millions of classified documents released by WikiLeaks as potential research materials. This paper examines trends in top-rated international relations journals that apparently cite or make reference to these materials in publications between 2010 and 2020. Findings indicate an increase in articles published by peer-reviewed journals that cite these materials, and a lack of professional consistency while doing so. This suggests the need for clear professional consensus within the research community surrounding these ethically questionable materials.

Table of contents

*Code and data are available at: https://github.com/ThomasWilliamFox/leaks_attribution_research.git; Replication on Social Science Reproduction Platform can be found at: https://doi.org/10.48152/ssrp-zvgn-8j27

1 Introduction

Researchers in various fields are guided by frameworks established by professional bodies and institutions while conducting research and publishing their work. Paleontologists and art historians for instance put considerable emphasis on the provenance of fossils and artworks while considering whether or not to engage with these sources (Darnton 2022). These procedures aim to ensure that ethical practices are involved in the procurement of these objects and to verify that these sources of information are trustworthy and authentic (Darnton 2022). As millions of classified government documents are leaked via digital platforms like WikiLeaks, researchers in political science and international relations are given little professional guidance surrounding the temptation to use these sources of information in their research (Darnton 2022).

This paper follows a reproduction of Darnton's 2022 paper, "The Provenance Problem: Research Methods and Ethics in the Age of WikiLeaks" by examining its replication data made available through the Harvard Dataverse (Darnton 2021). The analysis of the data was carried out with the programming language `R` (R Core Team 2023). Various packages were used in the processing, cleaning, and presentation of this information, including `tidyverse` (Wickham et al. 2019), `knitr` (Xie 2023), `here` (Müller 2020), `dplyr` (Wickham et al. 2023), and `janitor` (Firke 2023).

This analysis shows that influential, editorial, non-peer reviewed international relations journals may have set a precedent of publishing articles apparently referencing leaked classified materials in the years following the "Iraq War Logs" leak in 2010 (Darnton 2022). The paper's estimand is that as the numbers of articles published by these journals declined, peer-reviewed academic journals began publishing articles apparently referencing leaked classified documents with increasing frequency. These articles are inconsistently cited, provide little justification for using such materials, and rarely discuss the implications of these practices. A framework similar to that discussed by Darnton (2022) is supported by these findings as a potential means of encouraging professional reflexivity and consistency in academic and professional research.

This remainder of this paper contains an outline of the data examined, an analysis of the data, and a discussion of the findings. Section 2 outlines the analysis data made available by Darnton (Darnton 2021). Section 2.1 examines the TRIP-20 journals and their ratings Maliniak, Peterson, and Tierney (2012). Section 2.2 outlines the data-set built by Darnton, consisting of information about articles that apparently use classified materials (Darnton 2021). An exploration and analysis of the data is found in Section 3. The paper concludes with a discussion surrounding the paper's findings, including potential weaknesses and further research Section 4.

2 Data

2.1 TRIP Survey International Relations Journals (2011)

The 2011 Teaching, Research, and International Policy (TRIP) faculty survey was published in 2012 as part of the TRIP Project at The Institute for the Theory and Practice of International Relations at the College of William & Mary in Williamsburg, Virginia (Maliniak, Peterson, and Tierney 2012). The survey, collected responses from 3464 political science, international relations, and foreign policy professionals across 20 countries (Maliniak, Peterson, and Tierney 2012). The table on page 52 of the publication is a world-wide aggregate of responses on two questions surrounding which journals have the "greatest influence on the way IR scholars think about international relations" (Maliniak, Peterson, and Tierney 2012). The top 20 journals from the TRIP survey were used the basis for Darnton's analysis data set of articles apparently referencing classified documents (Darnton 2022). Table 1, is built using data found within Darnton's replication data set (Darnton 2021). The data consists of the names of the 20 journals, their short form abbreviations used by Darnton, as well as their ranking in the TRIP survey, with 1 being most influential and 20 being the least.

Table 1: Names, Shortform, and TRIP-20 ratings of the 20 Journals

Journal Name	Short Form	TRIP-20 Rating
International Organization	IO	1
International Studies Quarterly	ISQ	2
International Security	IS	3
Foreign Affairs	FA	4
American Political Science Review	APSR	5
World Politics	WP	6
European Journal of International Relations	EJIR	7
Journal of Conflict Resolution	JCR	8
Foreign Policy	FP	9
Review of International Studies	RIS	10
Millennium: Journal of International Studies	MIL	11
American Journal of Political Science	AJPS	12
International Affairs	IA	13
Security Studies	SS	14
Review of International Political Economy	RIPE	15
Journal of Peace Research	JPR	16
International Studies Review	ISR	17
International Relations	IR	18
Comparative Politics	CP	19
Global Governance	GG	20

2.2 Articles Apparently Referencing Materials From WikiLeaks (2010 - 2020)

A data set containing information related to articles apparently referencing classified materials released by WikiLeaks was constructed by Darnton as part of their paper (Darnton 2022). The terms, "Wikileaks", "cable", and a combination of the two were used to search for articles published between 2010 and 2020 in TRIP rated journals. Darnton used data bases such as JSTOR, Project Muse, and ProQuest, as well as publisher platforms including Sage, Wiley, and Oxford University Press to locate the articles contained in the data set (Darnton 2022). Of the 565 articles produced through their search, 397 were false positives (mentioned WikiLeaks organization or physical cables). Darnton read and itemized each of the 168 remaining articles to build their analysis data set (Darnton 2022).

Table 2 is built using Darnton's replication data (Darnton 2021). The data set contains 168 entries, each corresponding to an article in a TRIP rated journal that apparently references classified documents released by WikiLeaks. The "code" variable refers to Darnton's coding system, with "2" representing articles that refer to these materials via secondary sources, and "3" representing articles that quote, cite, or paraphrase materials from WikiLeaks directly (Darnton 2022). The "Journal" and "Year" variables refer to the journal that published the article and the year in which it was published.

Table 2: Sample of Cleaned Academic Publication Data

Code	Journal	Year	Mentions 'Classified'	No Source	Mentions 'Leak'
3	APSR	2020	n	n	n
3	IA	2020	n	n	n
3	IA	2020	n	n	n
3	IS	2020	y	n	n
3	ISR	2020	n	n	n
3	SS	2020	y	n	y

Uses euphemism	WikiLeaks' cable	To-From-Found	Peer Reviewed	Cable Discussed	Cable(s) Cited
n	n	y	y	y	0
n	y	n	y	y	4
n	n	n	y	n	0
n	n	n	y	y	8
n	n	n	y	y	4
n	n	n	y	y	11

Various variables coded as "y" for yes and "n" for no are used in Darnton's data set. The "Classified" variable indicates whether the article makes reference to the classified nature of

the materials (Darnton 2022). "No Source" indicates that at least one referenced document is missing a citation. "Leak" indicates that the article mentions leaked nature of sources. "Uses Euphemism" refers to articles that use verbs other than leak such as "publish, or disseminate." "WikiLeaks' cable" indicates if an article refers to a document or cable as the possession of WikiLeaks rather than its true source. The "To-From-Found" variable describes whether citations include an "author (can be"embassy"), recipient (even "State"), and repository" (Darnton 2022). "Peer reviewed" indicates whether the journal is peer reviewed. "Cable Discussed" indicates whether an article discusses leaked US cables and the "Cables Cited" variable communicates how many individual leaked cables are cited.

Figure 1 displays all 168 articles published in TRIP rated journals from 2010-2020 that apparently cite or reference leaked documents from WikiLeaks directly or through secondary sources. The years with the greatest number of articles published are 2011 with 24, 2013 with 21, and 2016 with 20. The Journals that published the most articles according to the data set are International Affairs (IA) with 29, Foreign Policy (FP) with 25, and International Security (IS) with 16.

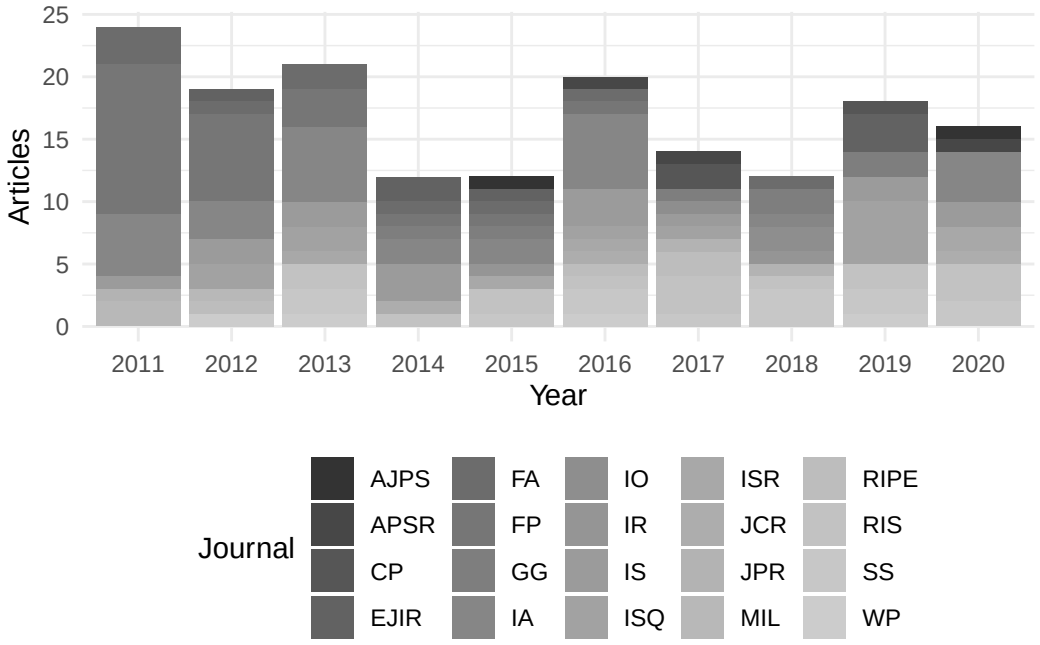

Figure 1: Articles in TRIP rated Journals apparently citing leaked classified material either directly or indirectly between 2010 and 2020

3 Results

To better understand the publication trends in these articles, this section will explore an analysis of the articles directly citing apparently classified documents. We will focus on differences in trends between peer-reviewed and non-peer-reviewed journals as well as citation attributes in the articles published by academic journals using these materials.

Figure 2 displays articles that apparently directly cite leaked classified materials in peer-reviewed journals. 81 articles in peer-reviewed TRIP rated journals apparently used such materials. The Journals with the greatest number of articles citing apparently leaked materials are International Affairs (IA) with 14, International Security (IS) with 10, as well as the Review of International Studies (RIS) and Security Studies (SS) with 9 each. This graph indicates an upward trend in publications in these journals from 2010 to 2020 with the greatest number of articles published in 2016, 2018, and 2019.

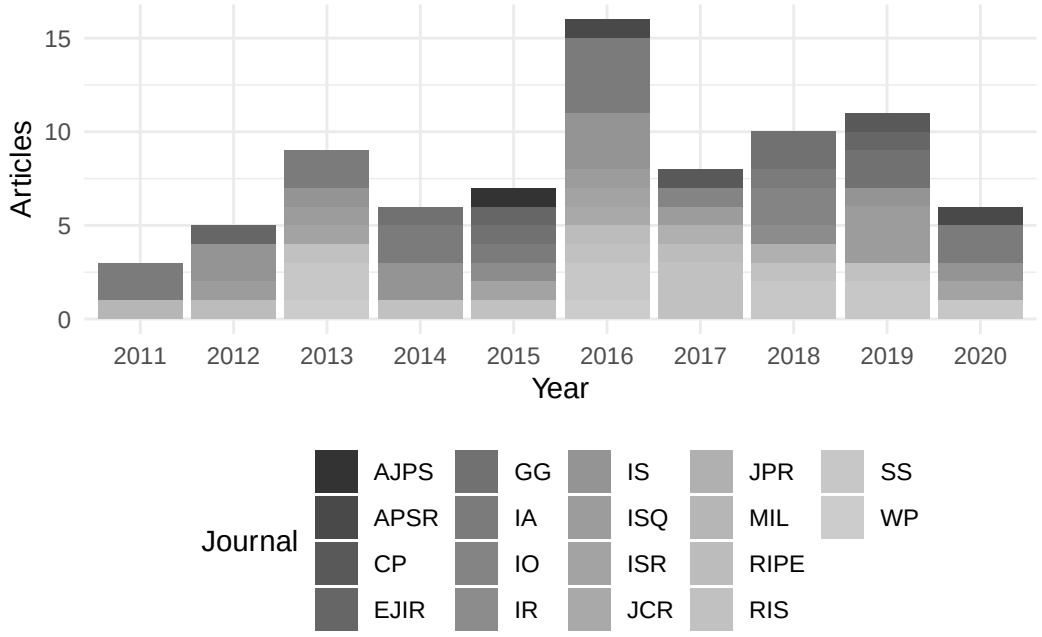

Figure 2: Articles in peer-reviewed TRIP rated Journals apparently citing leaked classified material directly between 2010 and 2020

Figure 3 shows the articles apparently directly citing classified materials published by Foreign Affairs magazine and Foreign Policy magazine, both non-academic, editorial journals (Darnton 2022). Foreign Affairs (FA) published 10 articles referencing such materials and Foreign Policy (FP) published 25. The majority of these articles were published from 2011 to 2013 where

Foreign Policy published 22 articles and Foreign Affairs published 6, after which a sharp decline in articles released by these publications can be seen.

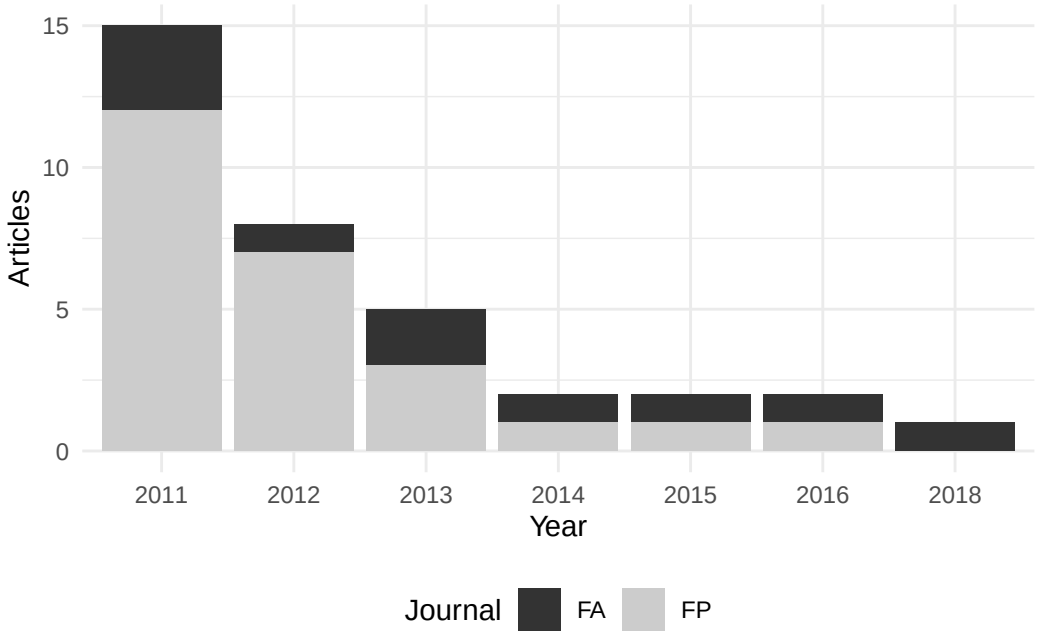

Figure 3: Articles in Foreign Affairs and Foreign Policy Magazines apparently citing leaked classified material directly between 2010 and 2020

Figure 4 shows the number of peer reviewed articles that either fail to mention the classified nature of the materials cited, omit any reference to the materials being leaked, or fail to properly cite the cable being referenced according to the to-from-found format outlined in Section 2.2. Any one of these attributes suggests that an article fails to communicate the true source of the utilized information. Of the articles in this group, 12 were published by International Affairs (IA), 8 by International Security, and 7 each by International Studies (RIS) and Security Studies (SS). Of the 64 articles that had one of the attributes outlined above, 33 articles had all three.

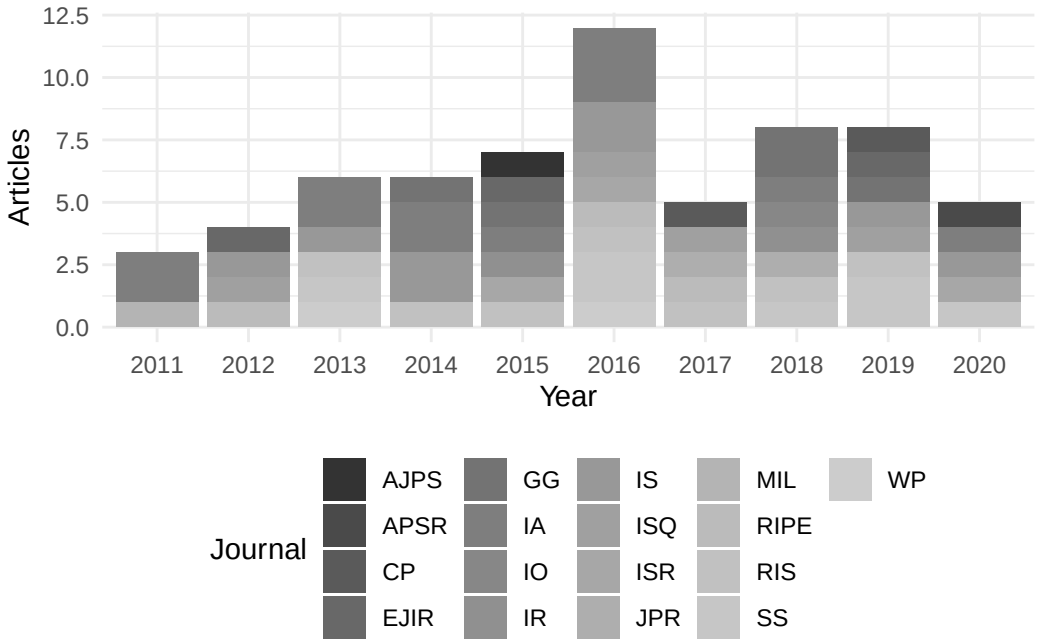

Figure 4: Articles in peer-reviewed TRIP rated Journals directely referencing apparently classified materials but failing to either communicate their apparently classified nature, mention their leaked orgin, or use a to-from-found citation format.

Figure 5 demonstrates peer-reviewed journal articles articles that either discuss the classified nature of the material, discuss the fact that the materials were leaked, or contain a "to-from-found" citation (see Section 2.2). Journals with the greatest number of articles that display one of these attributes are International Affairs (IA) with 7, International Security (IS) with 5, and 4 each by International Studies (RIS) and Security Studies (SS). It is worth noting that none of the 31 articles in this group display all three attributes.

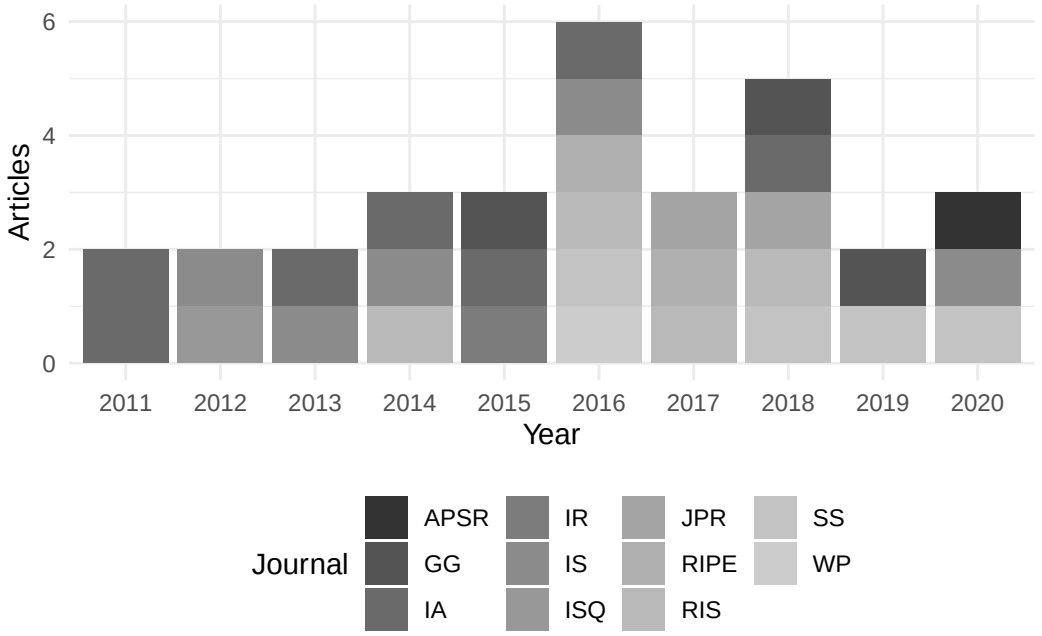

Figure 5: Articles in peer-reviewed TRIP rated Journals directly referencing apparently classified materials and either properly communicating their apparently classified nature, mentioning their leaked orgin, or using a to-from-found citation format.

These findings point to an increase in articles being published that apparently utilize these materials by peer-reviewed academic international relations journals. Although Foreign Affairs and Foreign Policy Magazine are not academic journals, they are both in the top ten most influential journals in the field of international relations (4 and 9 respectively) as outlined in the 2011 TRIP survey (Maliniak, Peterson, and Tierney 2012). The lack of consistency in disclosure regarding the nature of these documents in many of these articles obscures their source as well as the ethical implications surrounding their use. Many articles attempt to clearly communicate the source of these materials, but as Darnton points out, very few discuss the reasons for relying on them (Darnton 2022).

4 Discussion

This paper explores both the publication trends of articles in International Relations Journals apparently citing classified documents released via WikiLeaks between 2010 and 2020, as well as the inconsistencies of these practices in peer-reviewed academic Journals. By following a reproduction of Darnton's 2022 paper and utilizing their replication data set, a thorough exploration of these topics was made possible. The paper outlines the number of articles

published in peer-reviewed and non-peer-reviewed Journals over time, the number of articles failing to communicate the apparently classified nature of the cited information as well as articles that attempt to make this information clear.

4.1 Increase in Published Articles

As outlined in Figure 3, articles apparently citing classified documents published in non-peer-reviewed TRIP rated journals decreased sharply after a three year period ending in 2013 when 28 articles were published across both Foreign Affairs and Foreign Policy Magazines. The decreasing number of articles from 2011 to 2020 may suggest that these journals quickly published articles utilizing these sources when they were contextually relevant and moved on to other topics. In doing so, these types of journals may have set a precedent in the international relations publishing community that was followed by TRIP rated peer-reviewed academic journals.

Figure 2 demonstrates that each of the 18 TRIP rated peer-reviewed academic international relations Journals published at least one articles from 2010 to 2020 that cited apparently classified material. This shows an increase in articles published over time apart from outlier years in 2013 and 2016 (curiously surrounding U.S. election years). The lag in publishing following the Manning disclosures may be explained by lengthy research, review, and publishing times in academic journals. However, these delays, which can last upwards of 200 days after submission does not seem to account for the time discrepancy in publishing between peer-reviewed and non-peer-reviewed Journals (Powell 2016).

4.2 A Case for Reflexive Frameworks

4.3 Weaknesses and Next Steps

As this paper has followed a reproduction of Darnton's 2022 paper "The Provenance Problem: Research Methods and Ethics in the Age of WikiLeaks", its provided replication data set has been the basis for the data and analysis portions of this paper. This data set, as outlined in Section 2.2, was built by Darnton by manually searching various databases for articles containing keywords. Following this step, Darnton manually recorded all relevant attributes by searching and reading each article that directly cited the apparently classified materials. The various steps and tools used in manually searching and recording the data may produce some measure of error in the data set.

To conduct a more thorough replication, rebuilding Darnton's data set may be required in order to verify all articles and attributes, although this would require considerable time and resources. Broadening the scope of interest to collect more robust data may also be beneficial in including and comparing publication trends in Journals of other disciplines. Further research regarding the ethics and handling of classified documents in other fields of academic

research would be beneficial in exploring alternative professional frameworks. An investigation of historical document leaks and any prevalent publication trends may also be helpful in this regard. Future investigation could also include an analysis of any political, financial, or institutional interests held by publishers and the possible effect this may have on their willingness to publish these sorts of articles.

References

Alexander, Rohan. 2023. *Telling Stories with Data*. Chapman; Hall/CRC. https://tellingstorieswithdata.com.

Darnton, Christopher. 2021. *Replication Data for: The Provenance Problem: Research Methods and Ethics in the Age of WikiLeaks*. Harvard Dataverse. Dataset. https://doi.org/10.7910/DVN/SPLDTF.

———. 2022. "The Provenance Problem: Research Methods and Ethics in the Age of WikiLeaks." *American Political Science Review* 116 (3): 1110–25. https://doi.org/10.1017/S0003055421001374.

Firke, Sam. 2023. *Janitor: Simple Tools for Examining and Cleaning Dirty Data*. https://CRAN.R-project.org/package=janitor.

Maliniak, Daniel, Susan Peterson, and Michael Tierney. 2012. *TRIP Around the World: Teaching, Research, and Policy Views of International Relations Faculty in 20 Countries*. Williamsburg, VA: The College of William & Mary. https://www.wm.edu/ offices/itpir/_documents/trip/trip_around_the_world_2011.pdf.

Müller, Kirill. 2020. *Here: A Simpler Way to Find Your Files*. https://CRAN.R-project.org/package=here.

Powell, Kendall. 2016. "Does It Take Too Long to Publish Research?" *Nature* 530: 148–51. https://doi.org/10.1038/530148a.

R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

Wickham, Hadley, Romain François, Lionel Henry, Kirill Müller, and Davis Vaughan. 2023. *Dplyr: A Grammar of Data Manipulation*. https://CRAN.R-project.org/package=dplyr.

Xie, Yihui. 2023. *Knitr: A General-Purpose Package for Dynamic Report Generation in r*. https://yihui.org/knitr/.