

Seán Fobbe



Jonathan Gray/Wikimedia Commons

Open Data, Open Source und
Digitale Repositorien

Open Data, Open Source und Digitale Repositorien

Seán Fobbe

Legal:Inter:Faces — ZiF Arbeitsgemeinschaft

29. September 2022

Über diese Präsentation

Umfang

- Präsentation
- Source Code
- Vertiefungshinweise

Online verfügbar ca. 1 Woche nach Vortrag

Direkt www.doi.org/10.5281/zenodo.7111801

Blog www.seanfobbe.de

Drei Thesen

- 1 Nicht reproduzierbare Ergebnisse sind wissenschaftlich wertlos
- 2 Nur echte Beispiele können die empirischen Rechtswissenschaften mit Leben füllen
- 3 Open Access wirkt

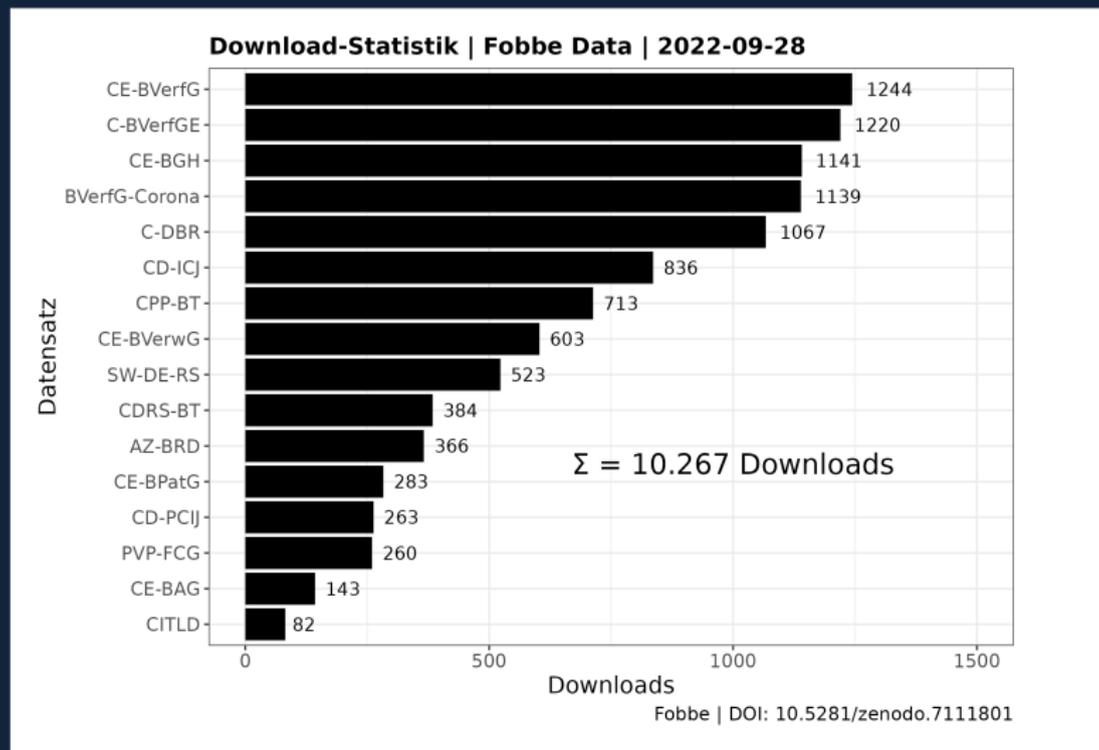
Reproduzierbarkeit in der Wissenschaft

- Wissenschaft fußt auf Wiederholbarkeit und Nachprüfbarkeit
- »An article about computational results is advertising, not scholarship« (Buckheit and Donoho 2010)
- Reproduzierbarkeit braucht Daten, Code und ggf. zusätzliche Angaben
- »Replication Crisis« in den Gesellschaftswissenschaften (z.B. Psychologie, Jura)

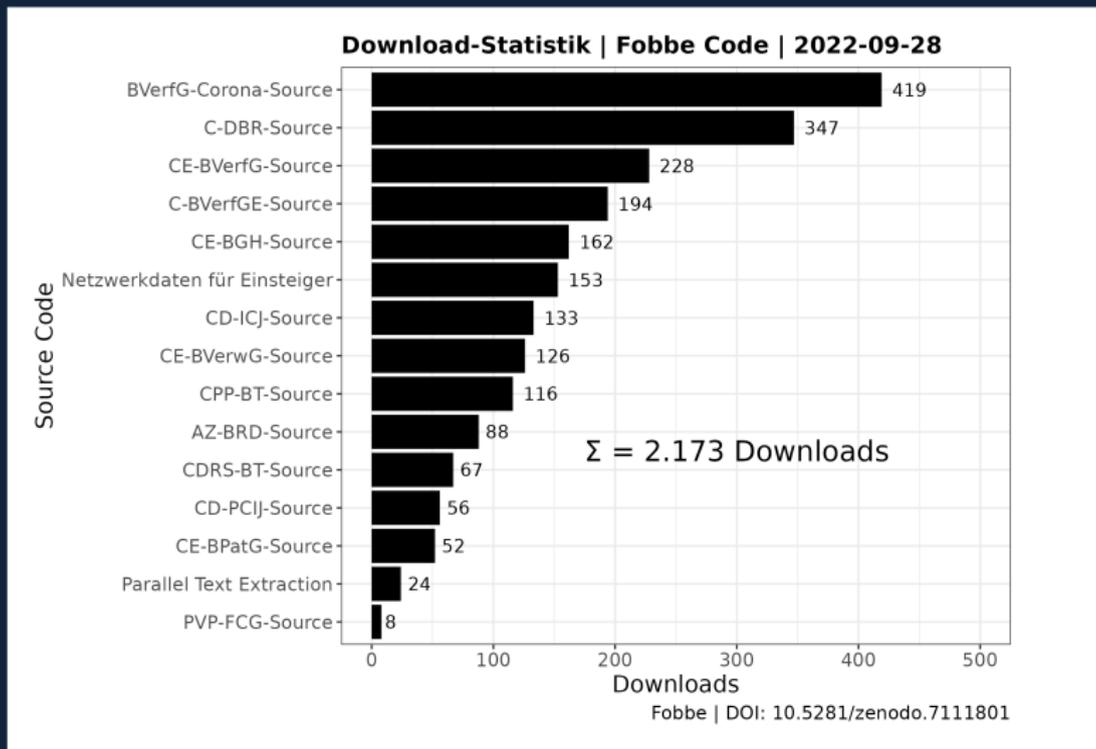
Notwendigkeit von echten Beispielen

- »Jurist:innen können sich in alles einarbeiten«
⇒ Lebenslüge der Rechtswissenschaft
- Jura ist heutzutage fast reine Literaturwissenschaft
- Keine Empirie? Wie realitätsfremd darf Rechtswissenschaft sein?
- Wie sollen Jurist:innen empirische Forschung lernen?
 - Mangel an empirisch bewanderten Forscher:innen
 - Mangel an frei nutzbaren und gut erklärten Daten
 - Mangel an beispielhafter empirischer Forschung

Open Data wirkt



Open Source wirkt



- »Literate Programming« macht Code auch für Nicht-Expert:innen wertvoll und lesbar
- Ausführliche Beschreibung verleitet eher zum Download
- Erstellen einer »Community« auf Zenodo erlaubt die Zusammenfassung von Forschungsleistungen einer Person/Gruppe
- Digital Object Identifiers (DOI) lassen sich im Voraus reservieren
- Es gibt eine Übungs-Seite: <https://sandbox.zenodo.org/>

Beispiel: Literate Programming

```
## [1] FALSE
```

```
print(mode.debug.example)
```

```
## [1] 5
```

3.9 DPI for OCR

This is the resolution at which PDF files will be converted to TIFF during the OCR step. DPI values will significantly affect the quality of text output and file size. Higher DPI requires more RAM, means higher quality text and greater PDF file size. A value of 300 is recommended.

```
ocr.dpi <- config$ocr$dpi  
print(ocr.dpi)
```

```
## [1] 300
```

3.10 Frequency Tables: Ignored Variables

This is a character vector of variable names that will be ignored in the construction of frequency tables.

It is a good idea to add variables to this list that are unlikely to produce useful frequency tables. This is often the case for variables with a very large proportion of unique values. Use this option judiciously, as frequency tables are useful for detecting anomalies in the metadata.

```
freq.var.ignore <- config$freq$ignore  
print(freq.var.ignore)
```

```
## [1] "date" "doc_id" "text"
```

3.11 Set Download Timeout

```
options(timeout = config$download$timeout)
```

Beispiel: Zenodo-Eintrag

The screenshot shows the Zenodo interface for a dataset titled "[R] Source Code der Corona-Rechtsprechung des Bundesverfassungsgerichts (BVerfG-Corona-Source)". The page includes a search bar, navigation links for "Upload" and "Communities", and a user profile for "s.fobbe@posteo.de". The main content area features a title, author information (Fobbe, Sean), and a description of the dataset. It also displays statistics such as 1,531 views and 415 downloads, and mentions that the software is updated every 6 months. A "Communities" section lists the dataset's association with the "Coronavirus Disease Research Community - COVID-19".

zenodo Search Upload Communities s.fobbe@posteo.de

September 8, 2022 Software Open Access Edit

[R] Source Code der Corona-Rechtsprechung des Bundesverfassungsgerichts (BVerfG-Corona-Source)

Fobbe, Sean

Überblick

Dieses R-Skript lädt den **Corpus der Entscheidungen des Bundesverfassungsgerichts (CE- BVerfG)** herunter, untersucht ihn auf mit SARS-CoV-2 assoziiertem Vokabular und speichert relevante Entscheidungen. Es ist die Grundlage für den Datensatz **Corona-Rechtsprechung des Bundesverfassungsgerichts (BVerfG-Corona)**.

Alle mit diesem Skript erstellten Datensätze werden dauerhaft kostenlos und urheberrechtsfrei auf Zenodo, dem wissenschaftlichen Archiv des CERN, veröffentlicht. Alle Versionen sind mit einem persistenten Digital Object Identifier (DOI) versehen. Die neueste Version des Datensatzes ist immer über den Link der Concept DOI erreichbar: <https://doi.org/10.5281/zenodo.4459405>

Aktualisierung

Diese Software wird **ca. alle 6 Monate** aktualisiert. Benachrichtigungen über neue und aktualisierte Datensätze veröffentliche ich immer sofort auf Twitter unter [@FobbeSean](#).

NEU in Version 2022-08-24

- Vollständige Aktualisierung der Daten
- Diagramme sind deutlich überarbeitet und die Labels verschönert worden
- Umbenennung des run scripts und der Konfigurations-Datei

1,531 views 415 downloads See more details...

Indexed in OpenAIRE

Quelle: www.doi.org/10.5281/zenodo.7062768

Beispiel: Zenodo-Community

Upload Communities 👤 s.fobbe@posteo.de

Open Access Code and Analysis Data by Seán Fobbe

Recent uploads

September 8, 2022 (2022-09-24) Software Open Access View

[R] Source Code der Corona-Rechtsprechung des Bundesverfassungsgerichts (BVerfG-Corona-Source)

👤 Fobbe, Sean;

Überblick Dieses R-Skript lädt den Corpus der Entscheidungen des Bundesverfassungsgerichts (CE-BVerfG) herunter, untersucht ihn auf mit SARS-CoV-2 assoziiertem Vokabular und speichert relevante Entscheidungen. Es ist die Grundlage für den Datensatz Corona-Rechtsprechung des

Uploaded on September 8, 2022

4 more version(s) exist for this record

September 7, 2022 (2022-09-27) Software Open Access View

Source Code for the 'Corpus of Decisions: International Court of Justice' (CD-ICJ-Source)

👤 Fobbe, Sean;

Overview This code in the R Programming Language downloads and processes the full set of decisions and appended opinions rendered by the International Court of Justice (ICJ) as published on its website into a rich and structured human- and machine-readable data set. It is the basis of the C

Uploaded on September 7, 2022

Published in Journal of Empirical Legal Studies, vol. 19, issue 2, p. 491.

1 more version(s) exist for this record

September 7, 2022 (11.8) Software Open Access View

Source Code for the 'Corpus of Decisions: Permanent Court of International Justice' (CD-PCIJ-Source)

👤 New upload

Community



Open Access Code and Analysis Data by Seán Fobbe

A curated collection of code, analysis data and computational results by or with contributions of Seán Fobbe.

For my writings and other publications please see here: <https://zenodo.org/communities/sean-fobbe-publications/>

For my stand-alone open access data sets please see here: <https://zenodo.org/communities/sean-fobbe-data/>

The full list of my publications is available under my ORCID (0000-0003-3808-7730). Link: <https://orcid.org/0000-0003-3808-7730>

Published by:

Social Media Tips

- Daten/Code/Präsentationen usw. auf Social Media posten (z.B. Twitter/LinkedIn) — auch bei kleiner Follower-Zahl!
- Gute Hashtags wählen, z.B.
 - #Rstats
 - #opensource
 - #opendata
 - #creativecommons
 - #jurabubble
- Twitter hat eine »Schedule«-Funktion für asynchrone Veröffentlichungen
- Nach »academic twitter tips« googeln

Drei Thesen und eine Lösung

- 1** Nicht reproduzierbare Ergebnisse sind wissenschaftlich wertlos
⇒ Veröffentlichung von Daten und Source Code
- 2** Nur echte Beispiele können die empirischen Rechtswissenschaften mit Leben füllen
⇒ Veröffentlichung von Daten und Source Code
- 3** Open Access wirkt
⇒ Veröffentlichung von Daten und Source Code

- FOBBE, Seán. Introducing Twin Corpora of Decisions for the International Court of Justice (ICJ) and the Permanent Court of International Justice (PCIJ). *Journal of Empirical Legal Studies*, 2022, 19. Jg., Nr. 2, S. 491-524. <https://doi.org/10.1111/jels.12313>
- COUPETTE, Corinna; HARTUNG, Dirk. Sharing and Caring: Creating a Culture of Constructive Criticism in Computational Legal Studies. arXiv preprint arXiv:2205.01071, 2022. <https://arxiv.org/abs/2205.01071>
- HAMANN, Hanjo; HÜRLIMANN, Daniel. Open Access bei der Veröffentlichung rechtswissenschaftlicher Fachliteratur—was soll das?. *Rechtswissenschaft*, 2019, Sonderheft, S. 3-30. <https://www.nomos-elibrary.de/10.5771/9783748903659-9.pdf>

Vielen Dank für Ihre Aufmerksamkeit!

Web www.seanfobbe.de

Twitter [@FobbeSean](https://twitter.com/FobbeSean)

E-Mail fobbe-data@posteo.de

- Titelbild: »Open Data stickers« von Jonathan Gray, CC0,
https://commons.wikimedia.org/wiki/File:Open_Data_stickers.jpg