



SFB/Transregio 266

ACCOUNTING FOR  
TRANSPARENCY

Making Open Science Matter: Wie können wir die  
Wiederverwendbarkeit und Wirkung von Forschung steigern?  
ZBW Symposium: Open Science – Forschungstransparenz in den  
Wirtschaftswissenschaften

Joachim Gassen

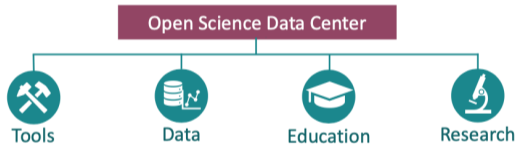
HU Berlin / TRR 266 Accounting for Transparency

April 27, 2023

Open Science is the practice of science in such a way that others can collaborate and contribute, where research data, lab notes and other research processes are freely available, under terms that enable reuse, redistribution and reproduction of the research and its underlying data and methods.

— FOSTER, <https://www.fosteropenscience.eu/>

## Was machen wir im Open Science Data Center vom TRR 266?



<https://www.accounting-for-transparency.de> <https://github.com/trr266>

- Open Science beschreibt einen “academic workflow”
- Wesentliche Auswirkungen auf
  - Arbeitsorganisation in Teams,
  - Verwendete Tools,
  - Datenmanagement,
  - Programmierung und
  - Dokumentation
- Idealer- und fast notwendigerweise projektbegleitend

## Was sind die Roadblocks auf der Angebotsseite?

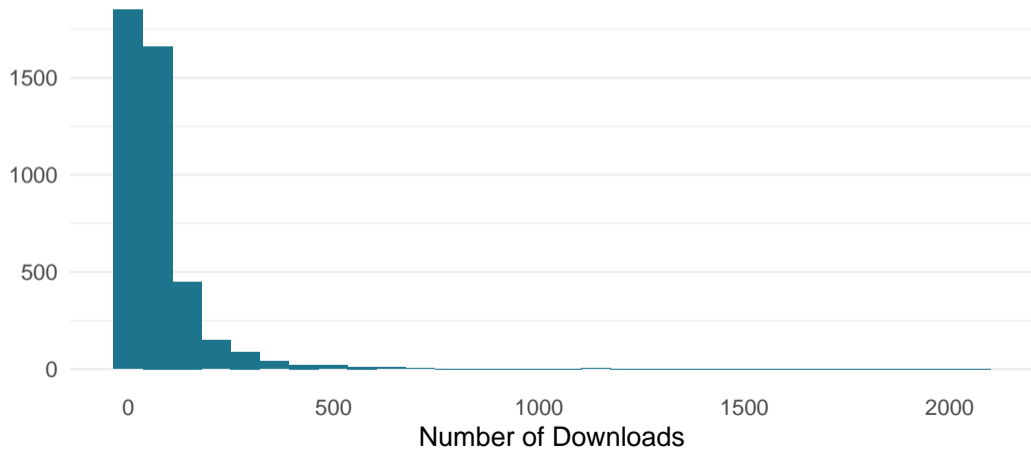
- Fehlendes Know-how
- Angst vor Fehlern
- Überhöhte Ansprüche
- Hierarchien in Teams
- Mangelnde Maintenance
- Lizenzierungsprobleme
- Datenschutz
- (Schutz des “geistigen Eigentums”)

## Was sind die Roadblocks auf der Nachfrageseite?

- Großes Interesse an wiederverwertbaren Daten und Tools aber eher geringes Interesse an spezifischen Projektmaterialien
- Mangelnde Interoperabilität der Daten
- Schwer auffindbare Inhalte
- Fehlendes Know-how
- Überhöhte Ansprüche (insb. hinsichtlich support)

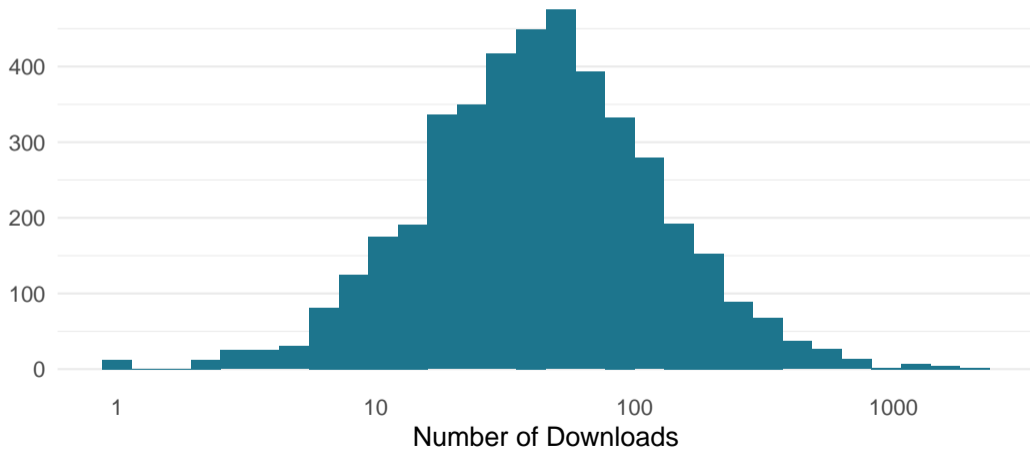
Wie steht es nun um die Wirkung von Open Science?

## Download numbers for AEA Replication Packages on OpenICPSR

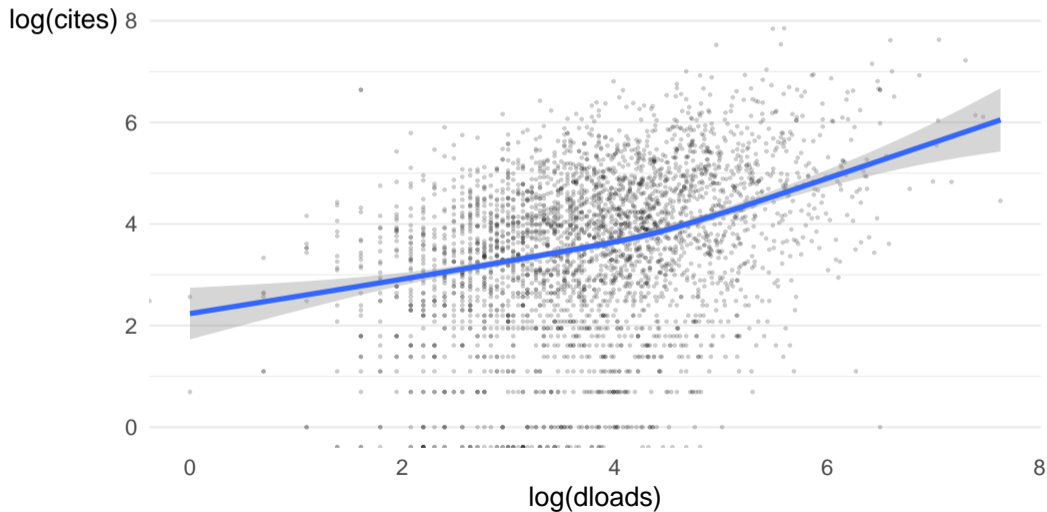




Replication packages werden im Schnitt etwa 75 mal runtergeladen



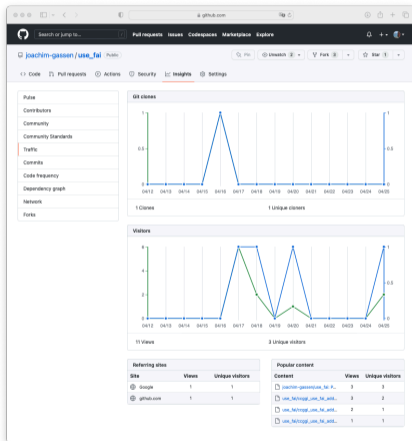
## Downloadzahlen korrelieren klar mit dem Impact des Papers



## Keine Daten ohne Regression ;-)

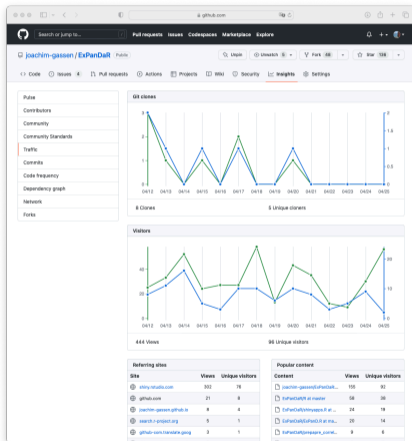
	log(Downloads)
log(Citations)	0.565*** [0.527, 0.602]
American Economic Journal: Applied Economics	0.176*** [0.057, 0.294]
American Economic Journal: Economic Policy	-0.295*** [-0.415, -0.175]
American Economic Journal: Microeconomics	-0.831*** [-1.075, -0.588]
American Economic Journal: Macroeconomics	-0.101* [-0.217, 0.014]
Journal of Economic Perspectives	-0.725*** [-0.946, -0.505]
Journal of Economic Literature	-0.815*** [-1.182, -0.449]
N	3,307
Adj. R <sup>2</sup>	0.485
Standard Errors	Clustered by Publication Year
Fixed Effects	Publication Year

# Persönliche Mikroevidenz: Ein Replication Package



Replication code for Cascino et al. (TAR, 2021)

## Zum Vergleich: Ein (nieschiges) R Package

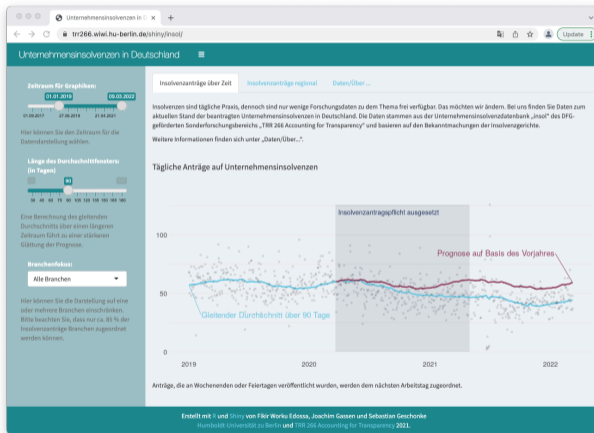


<https://github.com/joachim-gassen/expandar>

# Ein "Role Model"

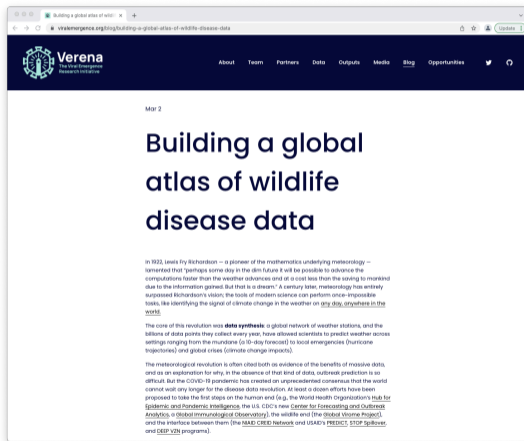
The screenshot shows the GitHub profile page for Vincent Arel-Bundock. The profile includes a circular avatar of a LEGO minifigure, the name "Vincent Arel-Bundock", and the bio "Political science prof". It also displays "494 followers - 4 following", location "Montreal, Canada", and website "http://varelbundock.com". The "Achievements" section shows several badges. The "Popular repositories" section lists: "modelsummary" (716 stars, 61 forks), "countrycode" (310 stars, 76 forks), "marginaleffects" (231 stars, 25 forks), "Rdatasets" (215 stars, 425 forks), "WDI" (177 stars, 68 forks), and "rethinking2" (66 stars, 22 forks). A "Contributions" calendar shows 4,204 contributions in the last year, with a grid of green squares indicating activity from May to April across the days of the week.

<https://github.com/vincentarelbundock>



<https://www.accounting-for-transparency.de/insol>

## Ein Beispiel aus einen anderen Feld



<https://www.viralemergence.org/blog/building-a-global-atlas-of-wildlife-disease-data>



Wieviele Unternehmen gibt es in Deutschland?

- Open Science is here to stay: Reproduzier- und Replizierbarkeit von Forschung sind zentrale Bestandteile des wissenschaftlichen Fortschritts
- Aber wir sollten dort nicht stoppen: Wirklichen Impact entfaltet Open Science immer dann, wenn andere auf den Ergebnissen und Methoden bisheriger Forschung aufbauen können
- Open Collaboration bringt einen wahren Performance Boost
- Sehr wichtig und leider schwierig: Dementsprechende Leistungen sind bei Karriereentscheidungen zu werten (Forschung oder Service?)



[https://github.com/joachim-gassen/zbw\\_oscience\\_talk](https://github.com/joachim-gassen/zbw_oscience_talk)

- Christensen, Freese and Miguel (2019): Transparent and Reproducible Social Science Research: <https://www.ucpress.edu/book/9780520296954/transparent-and-reproducible-social-science-research>
- Gentzkow and Shapiro (2014): Code and data for the social sciences: a practitioner's guide, <https://web.stanford.edu/~gentzkow/research/CodeAndData.pdf> (a nice and fun intro text)
- Wilson, Bryan, Cranston, Kitzes, Nederbragt and Teal (2017): Good enough practices in scientific computing, PLOS Computational Biology 13(6): 1-20, <https://doi.org/10.1371/journal.pcbi.1005510> (a pragmatic approach for the rest of us)