

# Forschungsdatenmanagement in einem interdisziplinären Forschungsprojekt

**Dr. Christian Riedel**

SFB 1294-Z03  
Information Infrastructure

Universität Potsdam  
Institut für Informatik  
und Computational Science



# Umfrage

Ich habe Daten/Skripte veröffentlicht (z.B. via GitHub)

Ich habe bisher noch keine Daten/Skripte veröffentlicht

# Umfrage

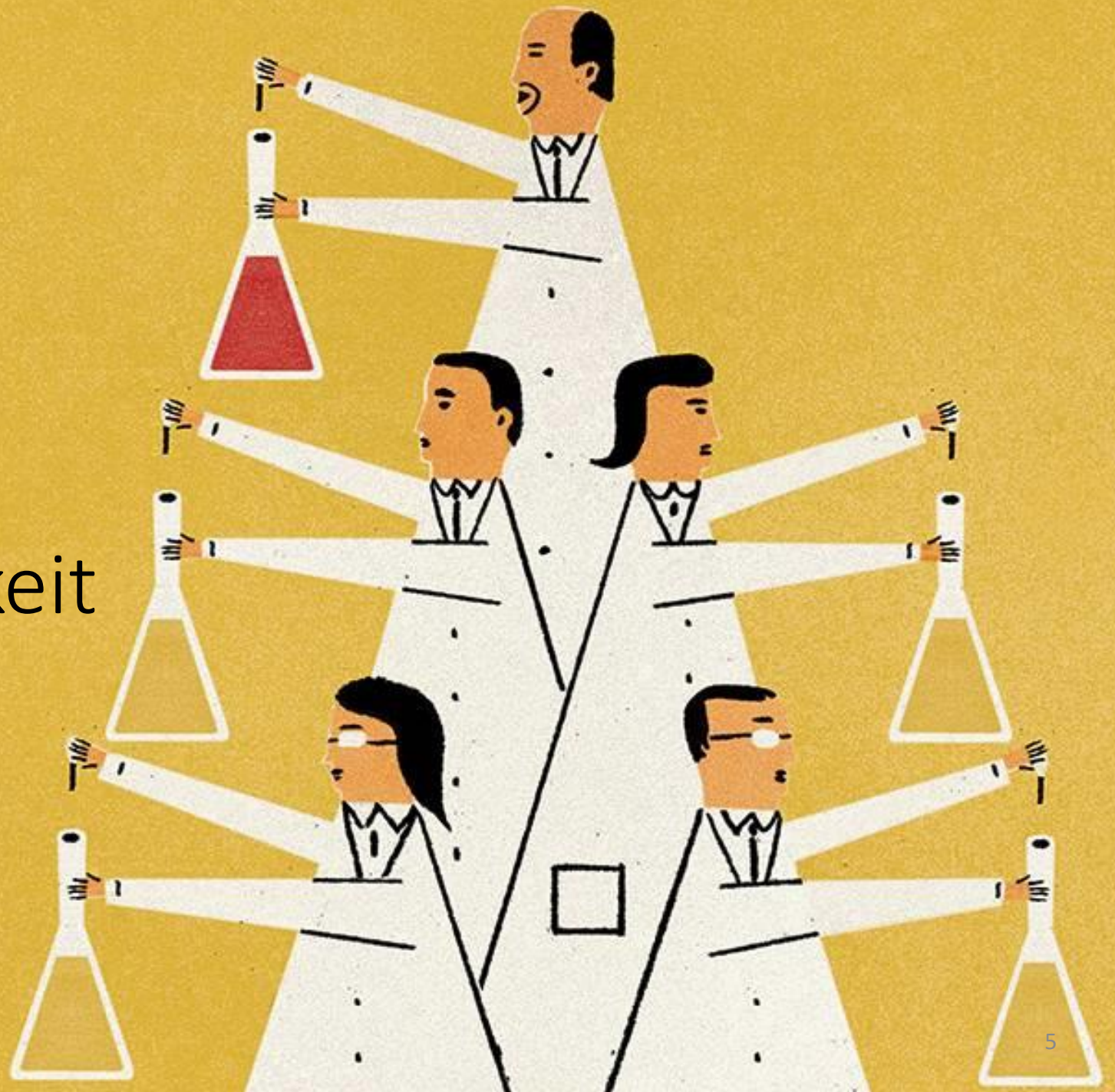
Ich habe Daten/Skripte von anderen genutzt (z.B. via GitHub)

Ich habe noch keine Daten/Skripte von anderen genutzt

Was ist Ihnen wichtig, wenn Sie die Daten/Skripte von anderen nutzen?



# Die Krise der Reproduzierbarkeit



# Die Reproduzierbarkeit von Forschungsergebnissen

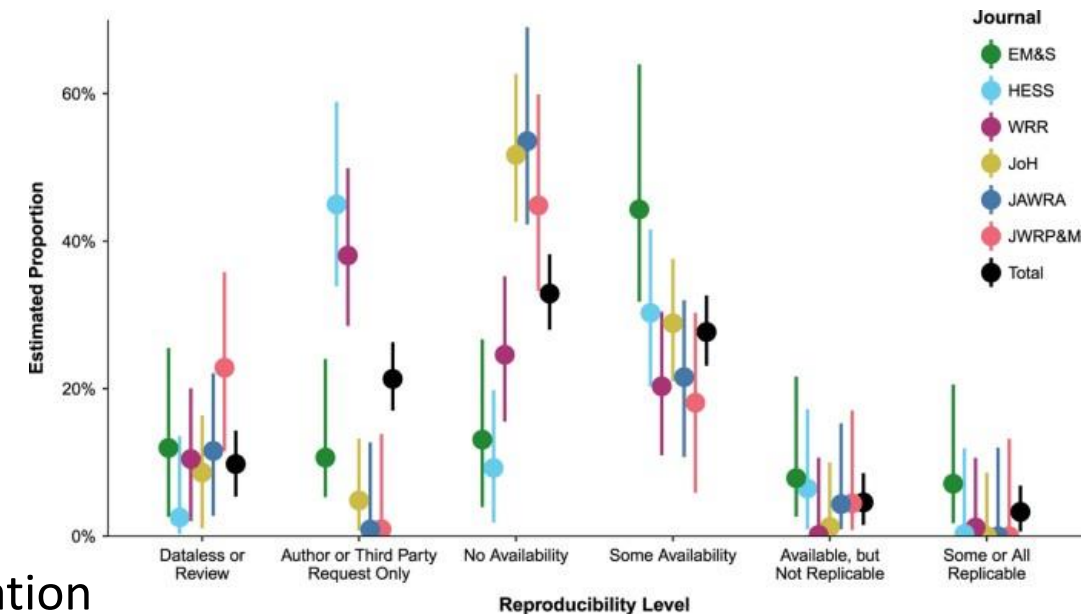
## 2019 Studie von Stagge et al.

Untersuchung von 360 Publikationen aus sechs Hydrologie & Water resource Journals aus 2017

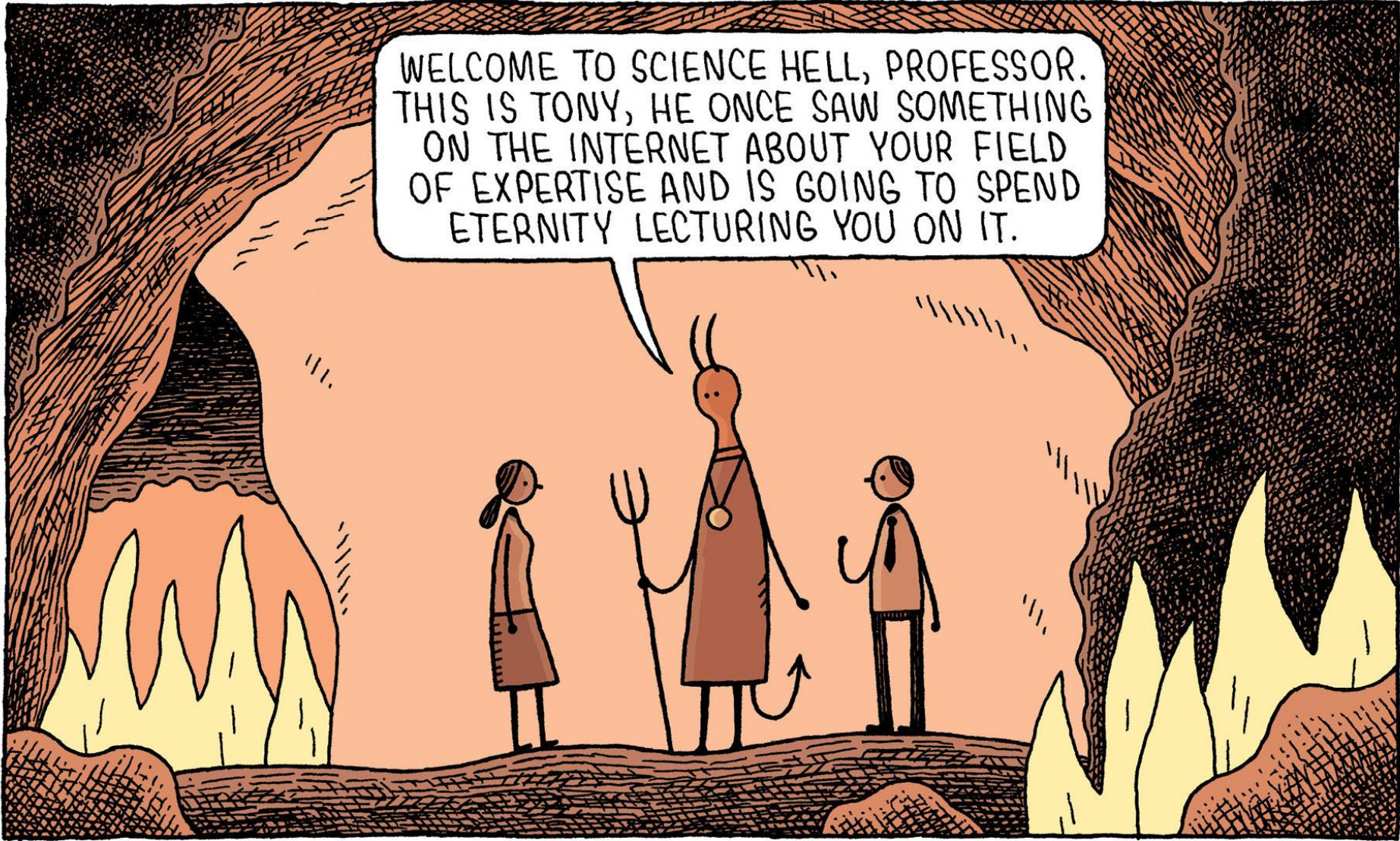
Bei nur 1.7% der untersuchten Artikel konnten die Ergebnisse reproduziert werden

**1.7% !**

Flaschenhals:  
Unvollständige Daten oder schlechte Dokumentation

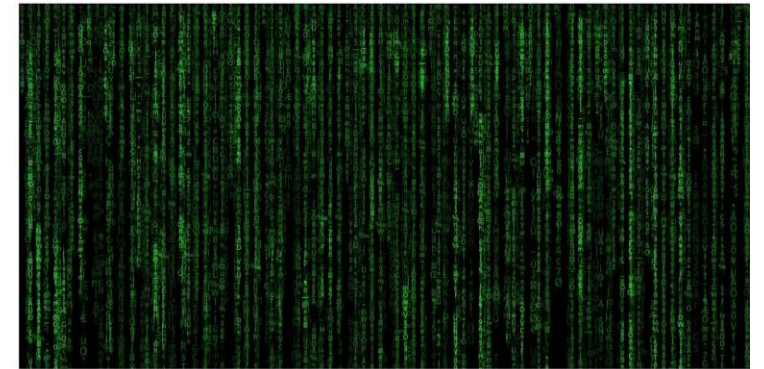
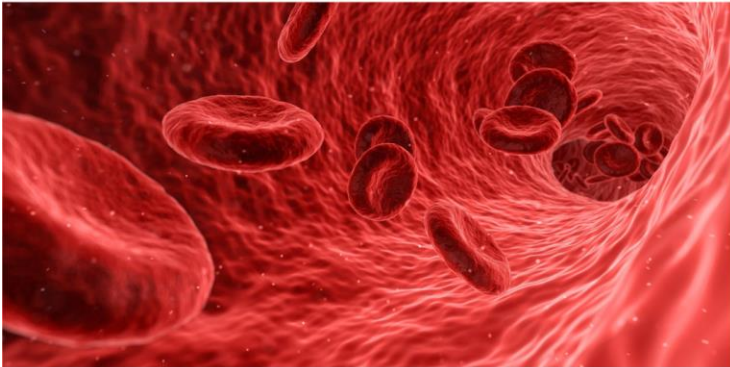






WELCOME TO SCIENCE HELL, PROFESSOR. THIS IS TONY, HE ONCE SAW SOMETHING ON THE INTERNET ABOUT YOUR FIELD OF EXPERTISE AND IS GOING TO SPEND ETERNITY LECTURING YOU ON IT.





## Entwicklung mathematischer Modelle und deren Anwendung

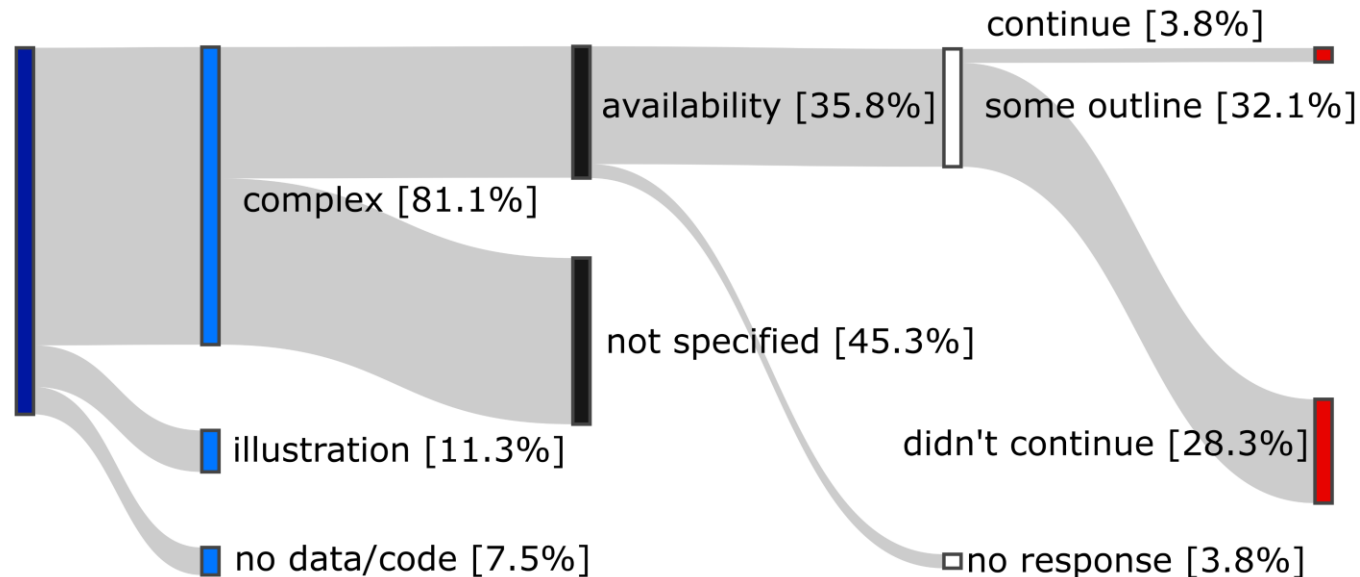


Verschiedene Disziplinen =

Verschiedene fachliche Hintergründe =

Verschiedene Daten

# Reproduzierbarkeit im SFB 1294 (2020)



- 53 Publikationen
- Nur 3,8 % der Publikationen hatte Daten, die auffindbar waren
- **Kein Experiment konnte reproduziert werden**

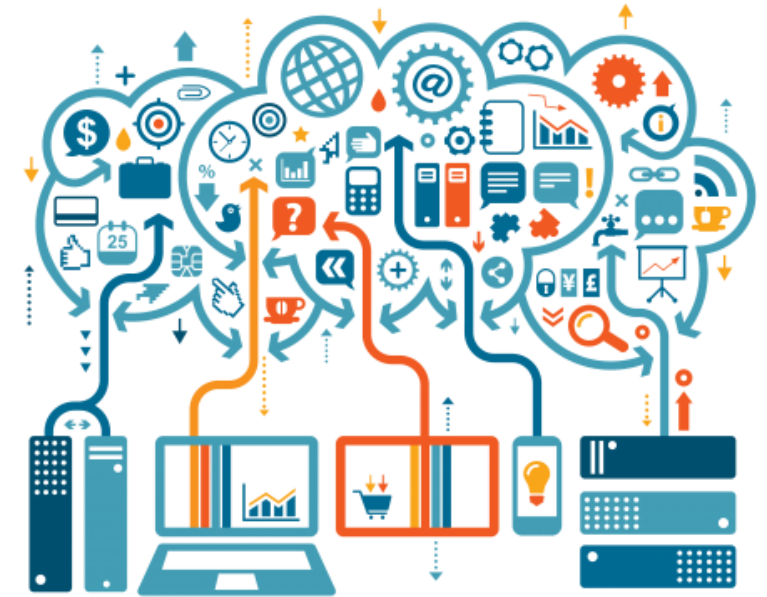
# Forschungsdatenmanagement

Alle Aktivitäten, die mit der

- Erhebung
- Speicherung
- Archivierung &
- Veröffentlichung

von **Forschungsdaten** zusammenhängen.

- Alle digitalen Daten, die während des Forschungsprozesses entstehen oder ihr Ergebnis sind



**Kann erstmal alles sein**

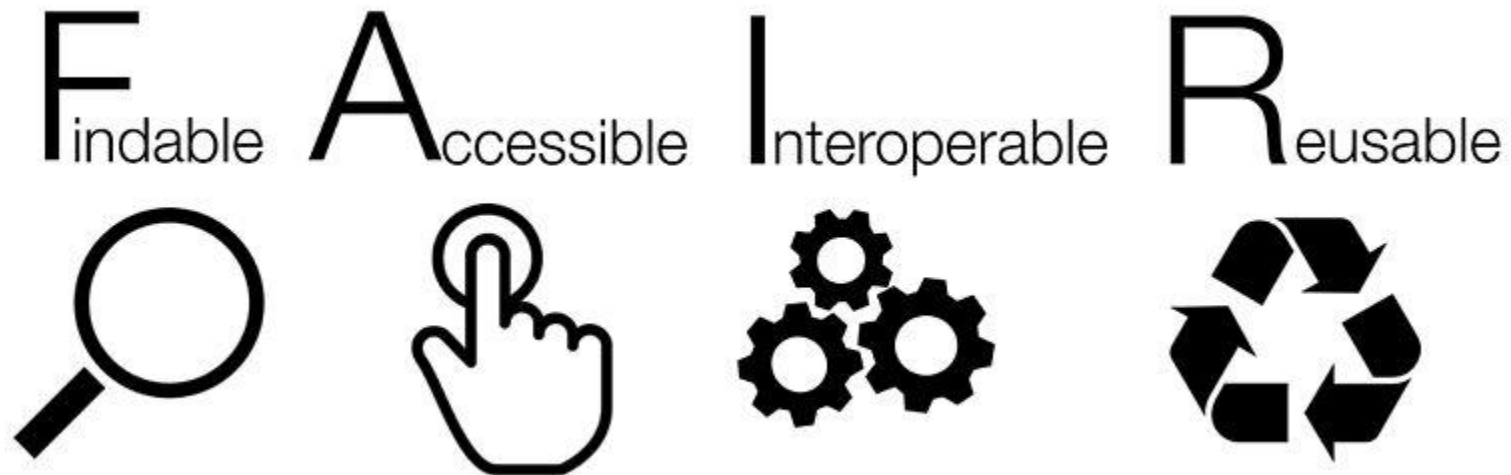


# Kreislauf von Forschungsdaten



Notwendige Schritte, um ein datenbasiertes Forschungsprojekt durchzuführen

# FAIR Prinzipien



G20 Staaten haben sich dazu bekannt, den Zugang zu öffentlich finanzierten Forschungsdaten, anhand der FAIR Prinzipien zu gewährleisten

[https://ec.europa.eu/commission/presscorner/detail/en/STATEMENT\\_16\\_2967](https://ec.europa.eu/commission/presscorner/detail/en/STATEMENT_16_2967)

# Handlungsempfehlungen - Policies

Können von Institutionen, Journales, Forschungsgemeinschaften etc. kommen

## Institution



- „Forschungsdaten sollen [...] so offen wie möglich zur Nachnutzung bereitgestellt werden.“
- „Forschungsdaten, [...] sind bei einer geeigneten Infrastruktur für **mindestens 10 Jahre zu archivieren.**“
- „Die Universität Potsdam erkennt die Aufbereitung von Forschungsdaten zur Nachnutzung sowie die Entwicklung nachnutzbarer wissenschaftlicher Software als **wissenschaftliche Leistungen an**“
- „Forschende Studierende und der wissenschaftliche Nachwuchs haben **Anspruch auf angemessene Information, Qualifizierung und Unterstützung** durch Lehrende und Betreuende“



# Umsetzung und Arbeit an der Basis

Reminder: Wir haben Guidelines, die versuchen, viele Forschungsfelder abzudecken und Forschende, die mit unterschiedlichsten Sachverhalten konfrontiert sind:

## Technische Anforderungen



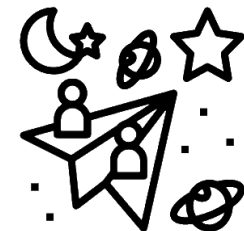
- Unterschiedliche Daten
- Unterschiedliche Software

## Erwartungen



- Datenschutz
- Abhängigkeiten  
(persönlich, Daten)

## Persönliche Realitäten



- Karriereplanung
- Kontrolle über  
Algorithmen abgeben

→ **Beeinflussen, wie sehr Forschende FDM-Policies umsetzen**

# Open Science & Kulturwandel in der Forschung

---

- **Vorteile (für Forschende und die Öffentlichkeit):**
  - Mehr Zitationen und größere Leserschaft
  - Wirtschaftlichkeit
  - Reproduzierbarkeit
  - Heuristik (Aufbauende Forschung vs. Doppelter Aufwand)
  - Wissenschaftliche Glaubwürdigkeit, Transparenz, Integrität
- **Schwierig:**
  - Akademische Realitäten, Unterschiedliche Erwartungen, Verschiedene Policies, etc. sind eine Herausforderung

**Wie lassen sich diese Prinzipien umsetzen?**

→ Es braucht einen Kulturwandel!



...und in der Praxis?



# 1. IT Infrastruktur



Plattformen zum  
kollaborativen Arbeiten  
& Wissensbasis

Wiki, Overleaf, Git,  
Box.UP storage,  
Datenarchiv, Data  
Management Plans



Erweiterung durch  
Einbindung von  
Industriestandards

Container  
Infrastruktur, High  
Performance  
Computing, CI/CD  
Pipelines

## 2. Vermittlung von Wissen



Unterstützung bei der  
Publikation von Daten durch  
Konsultationen und  
Workshops

Mitgliedschaft bei  
„The Carpentries“,  
Austausch mit  
Initiativen (MaRDI,  
NFDIxCS,  
Länderinitiativen)



Einstellung von  
wissenschaftlichen  
Hilfskräften zur Verbesserung  
der Reproduzierbarkeit

Unterstützung beim  
Testen und  
Dokumentieren,  
Umsetzung von good  
coding practices

# Fazit: Persönliche Erfahrungen



**IT Infrastruktur:**



Viele Gruppen im SFB nutzen die IT Infrastruktur



Mein Eindruck: Es gibt erhöhte Bereitschaft, die Plattformen zu nutzen



**Wissensvermittlung:**



Großer Bedarf an Workshops und Konsultationen



**Aber:**



Individuelle Verantwortungen können für Forschende sehr herausfordernd sein



Danke für die  
Aufmerksamkeit