

# 10 Thesen zu Open Data

Ulrich Dirnagl

Open Data Impact Award 2022, 7.11.2022

Dr. Evgeny Bobrov



# 1. 2022 war ein gutes Jahr für Open Science, und damit auch für Open Data

**Wissenschaftliches Publizieren  
als Grundlage und Gestaltungsfeld der  
Wissenschaftsbewertung**

**Open Science als Teil  
der Wissenschaftskultur**

Positionierung der Deutschen Forschungsgemeinschaft

Checkliste für einen angemessenen Umgang mit Forschungsdaten in DFG-Projekten

Dieser [Fragenkatalog](#) hilft Ihnen den Umgang mit Forschungsdaten in Ihrem Vorhaben zu planen und zu beschreiben.

**Leitfaden für qualitätsfördernde Aspekte  
in der Medizin und Biomedizin**



- Consider the **UNESCO recommendation on open science**<sup>38</sup>
- **Value activities associated with openness** (training, awareness raising, priority setting partnerships<sup>39</sup>, outreach)
- Value **outputs associated with openness** (FAIR data sets, pre-prints, open software, open code, translations)<sup>40</sup>

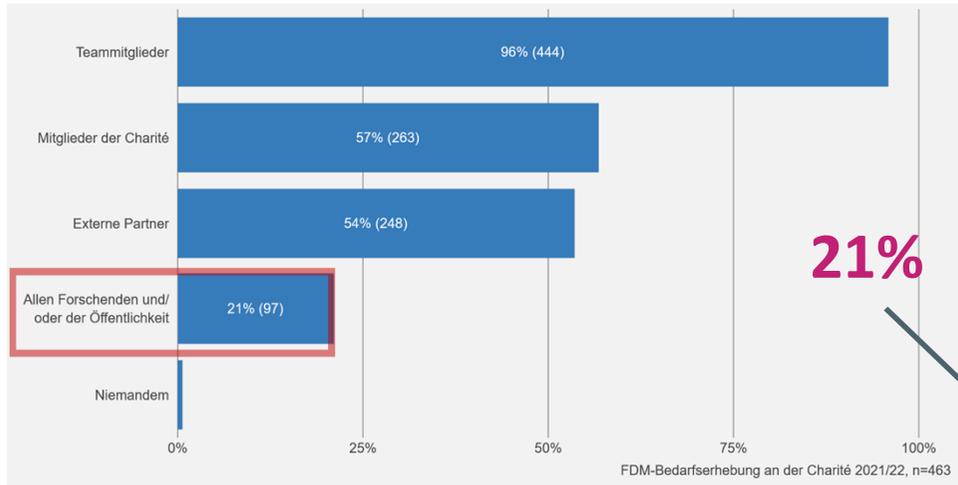
**Umgang mit Forschungsdaten**

Checkliste für Antragstellende zur Planung und zur Beschreibung des Umgangs mit Forschungsdaten in Forschungsvorhaben

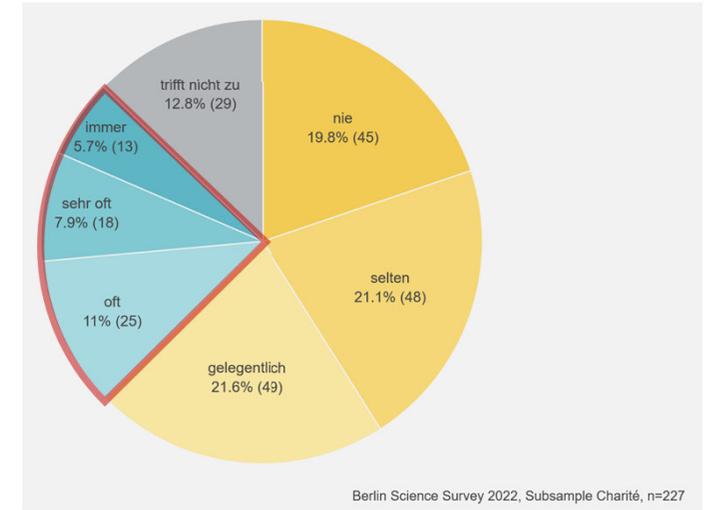
Coalition for Advancing Research Assessment <https://coara.eu/>

**DFG** Deutsche  
Forschungsgemeinschaft

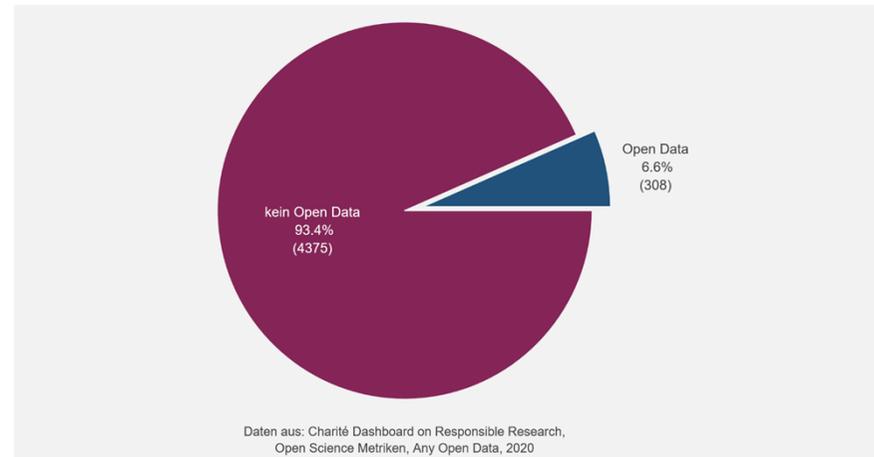
## 2. Teilen von Daten / OD ist noch die Ausnahme



21% ——— 25%  
7%



<https://quest-fdm-umfrage.charite.de/>



<https://quest-dashboard.charite.de>

# 3. Akademische Einrichtungen sollten Data/Code sharing (und FAIRness) monitoren und für ein Benchmarking öffentlich machen

## Charité Dashboard on Responsible Research

Charité has committed itself to establish, promote and maintain a research environment which enhances the robustness of research and the reproducibility of results ([Rethinking Health – Charité 2030](#)).

This dashboard gives an overview of several metrics of open and responsible research at the Charité (including the Berlin Institute of Health). For more detailed information on the methods used to calculate those metrics, the dataset underlying the metrics, or resources to improve your own research practices, click one of the following buttons.

This dashboard is a pilot that is still under development. More metrics will be added in the future.

For more detailed open access metrics you can visit the [Charité Open Access Dashboard](#) developed by the Charité Medical Library.

[See methods](#) [See resources](#) [See dataset](#)

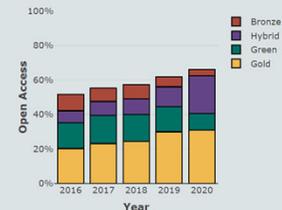
## Open Science

Show absolute numbers  Zoom in

### Open Access

62 %

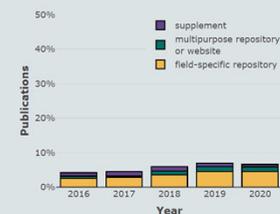
of publications were open access in 2020



### Any Open Data

7 %

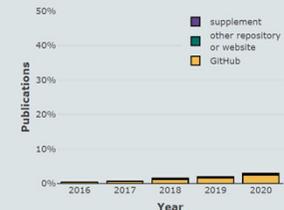
of publications mentioned sharing of data in 2020



### Any Open Code

132

publications mentioned sharing of code in 2020



### Preprints

254

preprints published in 2020



## FAIR assessment

Select repository type or repository

All repositories

### FAIR assessment by F-UJI

27 %

is the average FAIR score of research data objects in all repositories



# 4. Für das Monitoring wird eine Operationalisierung von Open Data benötigt



## Open Data Definition

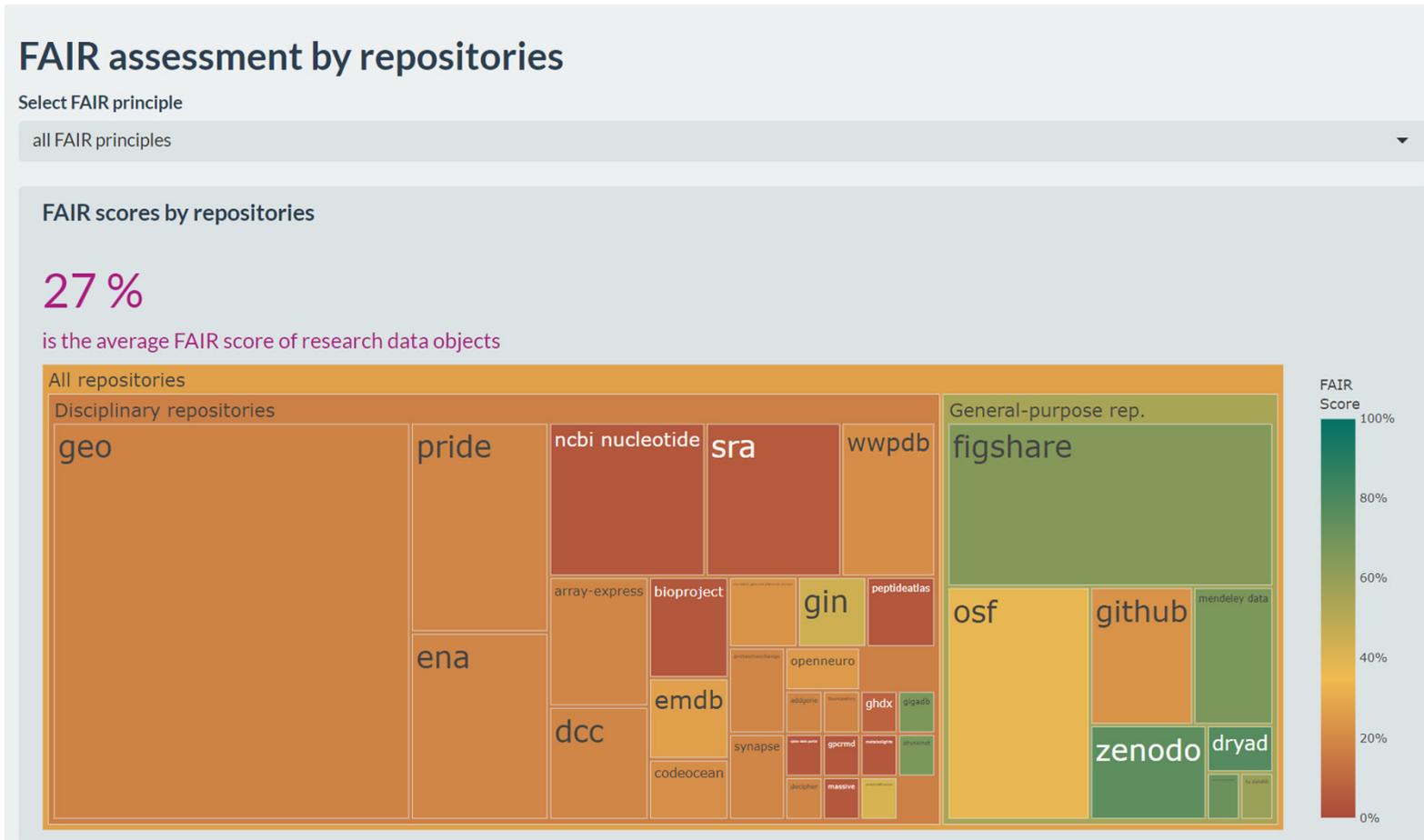
Open Data sind online kostenfrei zugängliche Daten, die verwendet, nachgenutzt und verbreitet werden können, mit allenfalls den Auflagen der Quellennennung und der Weitergabe unter gleichen Bedingungen (share-alike).



## Open Data Definition

Für die Zuweisung der Open-Data-Incentivierung als zusätzlicher Indikator der LOM (leistungsorientierte Mittelvergabe) Forschung an der Charité müssen die untenstehenden daten- und publikationsbezogenen Kriterien erfüllt sein. Allgemeine Kriterien für die LOM-Berechtigung der jeweiligen Person finden Sie auf der Internetseite unter Interne Forschungsförderung. Die Kriterien für die Open-Data-Incentivierung sind wie folgt (Stand 2022): Forschungsdaten wurden von Forschenden und Forschern der Charité frei zugänglich gemacht ODER die Daten wurden zugangsbeschränkt gestellt und erfüllen die folgenden Voraussetzungen: die Daten sind in einem externen Repository (z.B. Archiv, Datenbank, Register) abgelegt. es ist ein standardisierter Zugangsweg benannt, d.h. die Zugangsbedingungen, der Ablauf eines Antrags und die verantwortlichen Personen bzw. Stellen sind beschrieben. der Grund für den beschränkten Zugang wird genannt oder ist unmittelbar aus dem Personenbezug ersichtlich. der Zugang ist für das akademische Forschungs- /wissenschaftliche Umfeld / Wirtschaftsbereich / möglich. so-wissenschaftler / Wirtschaftsbereich / bei entsprechendem Fachwissen wird nicht zur Bedingung für die Bereitstellung der Daten gemacht. die Bereitstellung erfolgt kostenlos oder gegen Aufwandsentschädigung. Es kann sich um Rohdaten, Primärdaten oder Sekundärdaten (z.B. aus Analysen frei verfügbarer Datensätze, Meta-Analysen oder Health Technology Assessments) handeln; die Daten wurden somit eine analytische Replikation (Nachvollziehbarkeit der Analysemethoden) zumindest eines Teils der Ergebnisse der Studie ermöglichen; die Nutzung statistischer Zahlenwerte (Mittelwerte, Standardabweichungen, p-Werte etc.) nicht hierfür nicht aus. Daten wurden zu einer Archivpublikation gestellt, somit fallen für sich stehende Datensätze ohne Artikelbezug nicht darunter. Die Daten sind auch unabhängig von der Publikation auffindbar; somit sind Supplementary Materials nur zulässig, wenn sie in einem Repository (Archiv) abgelegt und auch über dieses Repository auffindbar sind. • In der Publikation wurde explizit auf die Datensätze hingewiesen; ein Verweis auf z.B. Supplementary Materials ohne weitere Erläuterung reicht nicht aus, ebenso wie der Verweis auf eine Datenbank ohne Nennung von Datensatz, Accession code oder genauen Suchbedingungen. Die Daten sind tatsächlich zugänglich und können zum Zeitpunkt der Überprüfung abgerufen werden (bei Daten unter Embargo muss dieses spätestens zum 31.7. ablaufen). Die Daten wurden in einem maschinenlesbaren Format gestellt, für Tabellen z.B. CSV-, Excel- oder Word-Daten, nicht jedoch PDFs oder Bildformaten. Im Folgenden fassen wir einige Aspekte zusammen, die nicht unter unsere Definition von Open Data fallen: Analyseskripte, Computerprogramme, Modelle und andere Methoden, Materialien oder Protokolle, auch wenn ihre Entwicklung Ziel der Forschungsprojekte / von anderer ihrer Darstellung / Hauptgegenstand der Publikation ist; wenn für die Entwicklung oder Validierung Daten erhoben und geteilt wurden, können diese jedoch unter die Open-Data-Definition fallen, sofern sie für sich stehend nachvollziehbar und nachnutzbar sind. Daten in Artikeltext selbst, sofern es sich nicht um eingebettete Tabellen handelt, die zugleich auch als für sich stehende digitale Objekte abgerufen werden können. Bildliche, audiovisuelle und andere Daten, die primär der Illustration dienen. Daten zu Fallreihen (case reports), sofern sie nicht in Repositorien aus der jeweiligen Fachdisziplin abgelegt wurden. Bei Systemic Reviews und Meta-Analysen: Listen mit Quellenangaben oder andere allgemeine Informationen zu den untersuchten Studien wie Erhebungsmethode oder Triebmedien. LOM-fähig sind dagegen neu aus der Originalarbeit zusammengestellte Datensätze, die eine nachvollziehbare / der Analyse sichernde, wie z.B. erkrankte Teilnehmer, oder statistische Werte) Daten, die nur auf Nachfrage und/oder bei Erfüllung von Voraussetzungen verfügbar sind, es sei denn, es handelt sich um personenbezogene oder anderweitig sensible Daten, die über einen standardisierten nachfragefrei verfügbar sind (darunter fällt z.B. nicht die Verfügbarkeit „upon request“ Daten aus Datenanfragen von Konsumenten („DataPods“), wenn unklar ist, ob die Autor:innen selbst einen Beitrag zum Pool geleistet haben. Daten, die für nur ein „private Link“ gestellt wird, so dass sie nicht im Repository gefunden werden können, sondern ausschließlich über die Publikation zugänglich sind. Daten, die vor dem betrachteten LOM-Zeitraum gestellt wurden (hierdurch soll sichergestellt werden, dass nur eine begrenzte Zahl von Artikeln für das Füllen eines bestimmten Datensatzes beibehalten werden kann).

## 5. Viele Repositorien sind ‚un-FAIR‘



## 6. Kuratierung von nachnutzbaren Daten frisst Ressourcen und erfordert Kompetenz – an beidem mangelt es häufig



[www.digitalbevaring.dk](http://www.digitalbevaring.dk)

## 7. Data sharing/OD muss gefördert/belohnt werden

Open data LOM („performance oriented funding“) allocated on institutional level („Einrichtungsebene“) to incentivize the set up of open data facilitating structures, knowledge base, support etc...



The screenshot shows the Charité application service website. The main heading is "Professorship Application Service" with a navigation menu including Home, Registration, Login, and FAQ. Below this is a section titled "Currently Vacant Professorships" with a table of open positions.

Description of Professorship	calls	notes	date due
W3 Universitätsprofessur für Hämatologie und Onkologie m. S. Lymphome/Multiples Myelom	<a href="#">call for applications</a>	<a href="#">application notes</a>	28.09.2017
W3 Universitätsprofessur für Hämatologie und Onkologie m. S. Medizinische/Internistische Onkologie	<a href="#">call for applications</a>	<a href="#">application notes</a>	28.09.2017

Further vacancies at Charité



“The Charité attaches great importance to transparent, replicable research and supports the objectives of Open Science (Open Access, Open Data). This includes the registration of studies in registries (clinicaltrials.gov, DRKS, etc.), the preregistration of studies, and the publication of negative and zero results. **How have you been pursuing these goals so far and what are your plans for the future?**”

## 8. Nicht nur das Teilen von Daten /OD muss gefördert/belohnt werden, sondern auch deren Nachnutzung

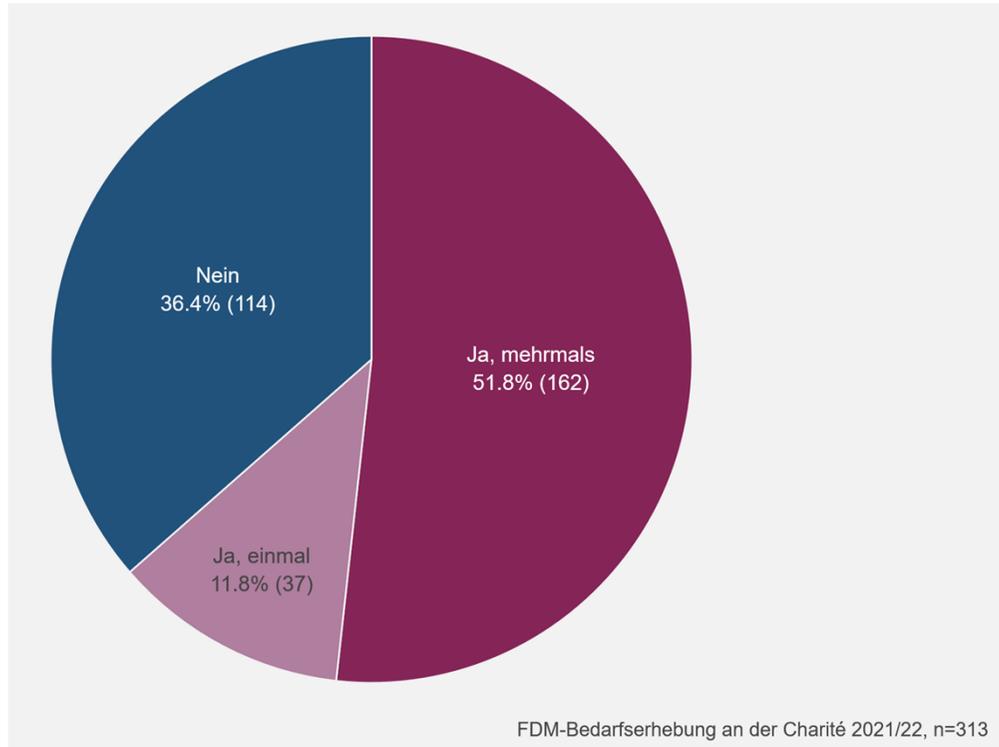


The QUEST 1,000 € Open Data Reuse Award  
*Make use of Open Data!*

<https://www.bihealth.org/de/translation/innovationstreiber/quest-center/ausschreibungen-und-preise/quest-ausschreibungen-und-preise/1000-eur-quest-preis-fuer-die-wiederverwendung-von-daten>

## 9. Datenschutz ist das größte Hindernis für OD (zumindest in Medizin)

*Haben Sie bereits einmal vorrangig aus Datenschutzgründen Forschungsdaten nicht geteilt?*



<https://quest-fdm-umfrage.charite.de/>

**64% haben Daten mindestens einmal aus Gründen des Datenschutzes nicht geteilt**

Schutz der eigenen Datennutzung mit Abstand auf Platz 2 (20%, n = 32)

## 10. Die Fördergeber sind die Treiber