

Wie FAIR sind unsere Forschungsdaten?

Das FAIR Data Dashboard der Charité – Universitätsmedizin Berlin



Evgeny Bobrov
evgeny.bobrov@bih-charite.de

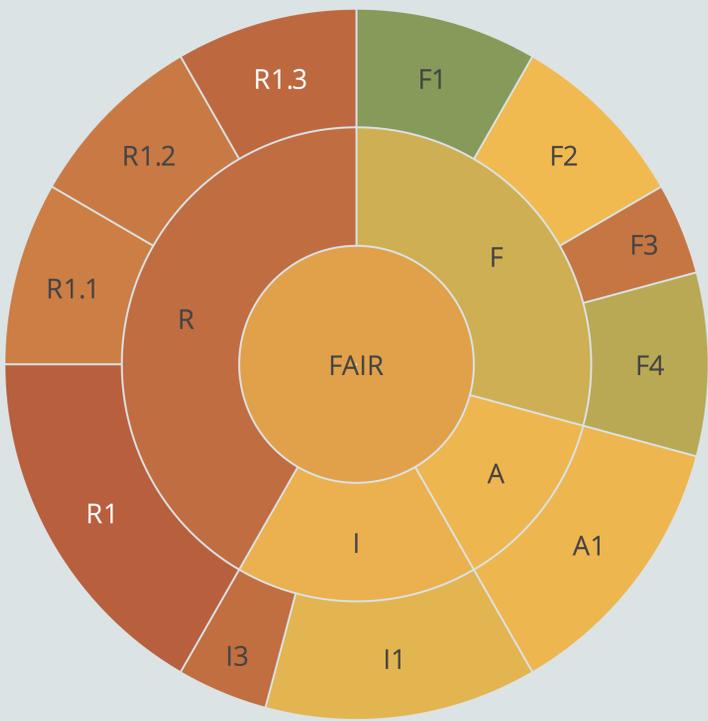
Maaïke Duine
maaïke.duine@open-access-berlin.de

Maxi Kindling
maxi.kindling@open-access-berlin.de

Jan Taubitz
jan.taubitz@charite.de

Wie es an der Charité – Universitätsmedizin Berlin um die Implementierung der FAIR-Prinzipien steht, wurde in einem von der Berlin University Alliance geförderten Projekt analysiert. Die Ergebnisse wurden ins bestehende Charité Dashboard on Responsible Research integriert.

Analysiert wurden Forschungsdaten, die gemeinsam mit im Jahr 2020 publizierten Artikeln öffentlich in Datenrepositorien geteilt wurden. Für jedes im Artikel zitierte Repository wurde ein Identifier (URL oder DOI) an das Screening-Tool F-UJI gesendet. F-UJI berechnet anhand von Metriken und Tests einen Score, der pro Datensatz die prozentuale Erfüllung der einzelnen FAIR-Prinzipien abbildet.

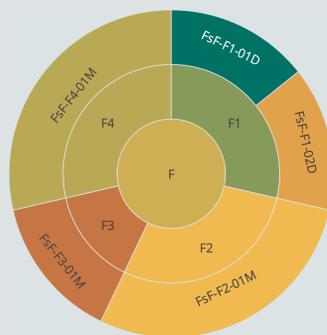


∅ FAIR Score aller Datensätze

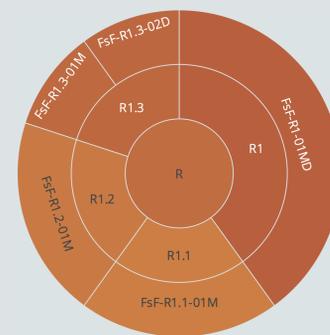
11 (der 15) FAIR-Prinzipien werden mit dem FAIR Assessment Tool F-UJI evaluiert

Die Ergebnisse der Screenings wurden u.a. auf Ebene der FAIR-Prinzipien und der Repositorien aggregiert.

Der durchschnittliche FAIR Score der 299 untersuchten Forschungsdatensätze beträgt 27%. Dabei ist die Auffindbarkeit (*Findability*) der Daten mit 45% recht gut sichergestellt. Die *Nachnutzbarkeit (Reusability)* erreicht aufgrund fachspezifischer Anforderungen dagegen nur einen durchschnittlichen Score von 11%.



∅ FAIR Score F-Prinzip



∅ FAIR Score R-Prinzip

Was ist FAIR?

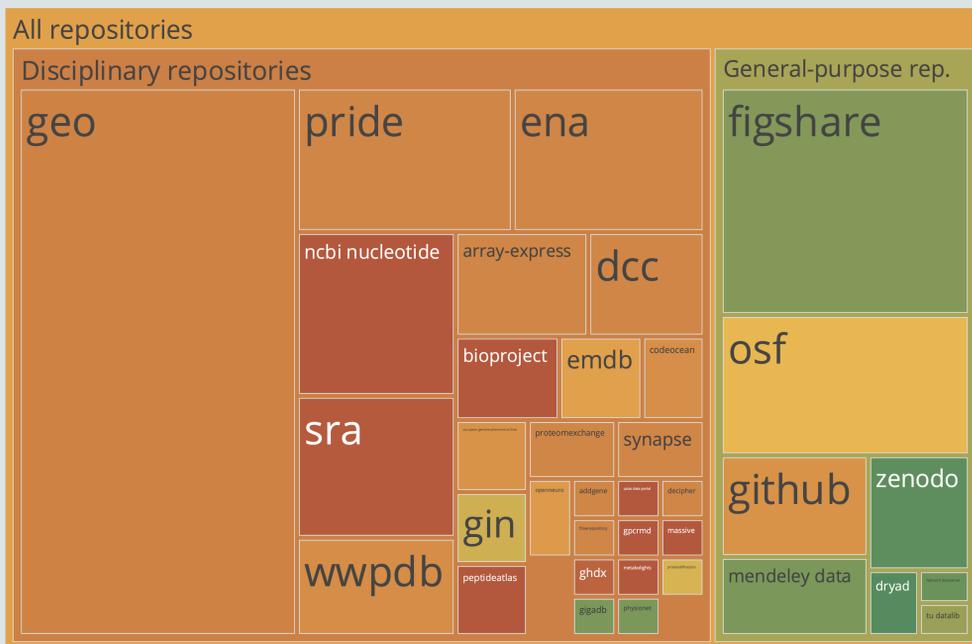
- 2016 veröffentlicht als FAIR Guiding Principles for scientific data management and stewardship
- Open-Data-Grundsätzen folgend sollen die FAIR-Prinzipien sicherstellen, dass Forschungsdaten *findable, accessible, interoperable* und *reusable* sind
- Der Fokus liegt auf der Maschinenlesbarkeit von (Meta-)Daten

QR-Code zum Dashboard



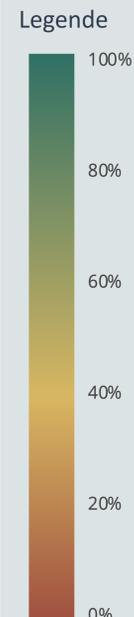
FAIRness wird zu einem großen Teil durch die Repositorien bestimmt

Allgemeine Repositorien – wie etwa Zenodo oder figshare – erzielen mit durchschnittlich 54% einen deutlich höheren FAIR Score als fachspezifische Repositorien – wie etwa das an der Charité intensiv genutzte NCBI Gene Expression Omnibus (geo). Diese Repositorien erreichen im Durchschnitt nur einen FAIR Score von 17%. Sie bieten jedoch den Vorteil fachspezifischer Metadaten.



∅ Fair Score nach Repositorien

FAIR Score



Dashboard

Visualisierung der Ergebnisse, methodische Beschreibung u. Veröffentlichung der aggregierten Daten



Analyse

Aggregation und statistische Analyse der Ergebnisse



High Level Flowchart

Eine Data Pipeline mit selbstentwickelten und nachgenutzten Tools führt zum FAIR Data Dashboard

Publikationsliste

Suche nach Charité-Publikationen in bibliographischen Datenbanken (WoS, Embase)



ODDPub

Text Mining der Artikel-Volltexte zur Identifikation von Data Statements



Numbat

Manuelles Screening der Artikel zur Validierung der Forschungsdaten-IDs



F-UJI

Automatisches FAIR Assessment

