

Trainingsdatensätze

Johannes Signer¹ (jsigner@uni-goettingen.de), Ulrike Schlägel²,
Daniel Tschink³, Juliane Röder⁴

¹Abteilung Wildtierwissenschaften, Georg-August-Universität Göttingen; ²Graduate School HIGRADE Helmholtz Centre for Environmental Research, Leipzig; ³GFBio e.V., Bremen; ⁴Stabsstelle Forschungsdatenmanagement, Philipps-Universität Marburg



Motivation

Für gute Lehre und Ausbildung von Nachwuchsforschenden sind Trainingsdatensätze essenziell. Sie bilden die Grundlage für die Vermittlung von Datenkompetenzen und Analyseverfahren.

Eigenschaften

Trainingsdatensätze sollten:

- **FAIR** (Findable, Accessible, Interoperable, Reusable) sein.
- frei verfügbar sein, mit einer entsprechenden Lizenz und offenen Datenformat.
- einen angemessenen Umfang haben.
- zitierbar sein.
- es ermöglichen leichtverständliche, aber interessante Fragestellungen zu bearbeiten.
- ausreichend dokumentiert sein.
- entweder "perfekt" oder Datensätze mit didaktischen Fehlern sein.

Was sind Trainingsdatensätze für uns:

- Datensätze, um alle Abschnitte des Datenlebenszyklus zu üben.
- Keine Datensätze, um Modelle zu trainieren (z.B. KI)

Arten von Trainingsdaten

Trainingsdatensätze können an unterschiedlichen Stellen im Datenlebenszyklus einsetzen.

Archivieren, Teilen und Nachnutzen von Daten.



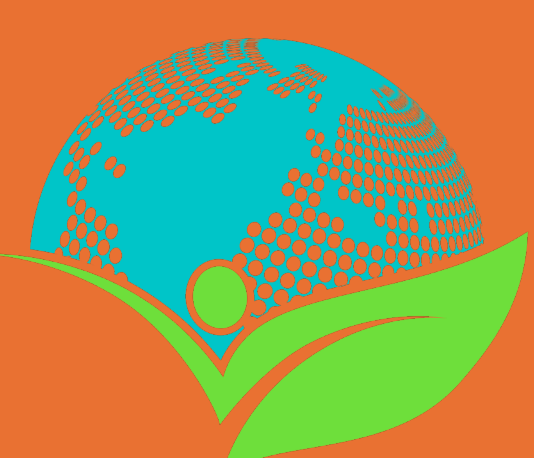
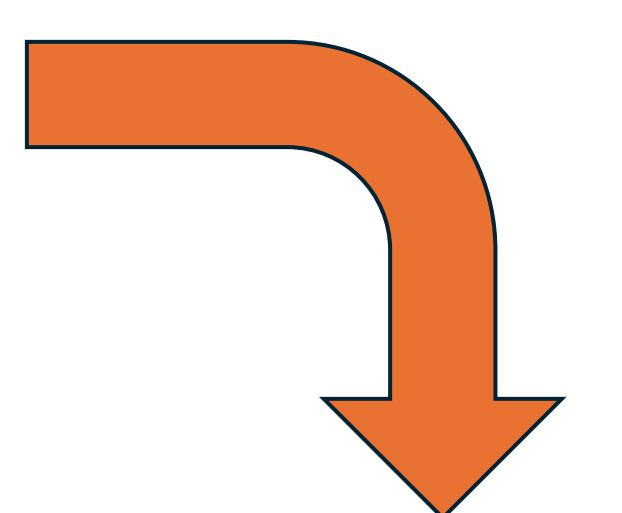
(Meta)Datenschema zum Aufnehmen von Daten.

Aufbereiten von Daten.

Analyse von Daten.

Ausblick

Haben Sie Interesse an Trainingsdatensätzen oder arbeiten Sie mit Trainingsdatensätzen? Dann treten Sie gerne mit uns in Kontakt (jsigner@uni-goettingen.de) und füllen Sie unsere **Umfrage** dazu aus.



NFDI 4
BIODIVERSITY

^aQuelle der Abbildung: <https://www.dsc-ub.de/beratungen.php>
(Zugriff 16.9.2024)



Poster (pdf)



Umfrage