

F.A.I.R. Data

Geosoftware II

Lara Sieksmeier, Amelie Luschnig

Gliederung

- Einleitung
- Die F.A.I.R.-Prinzipien
- Unterschied zwischen Open Data und F.A.I.R. Data
- Warum sollte man Daten F.A.I.R. machen?
- Bedeutung der F.A.I.R.-Prinzipien allgemein
- Bedeutung der F.A.I.R.-Prinzipien in der Wissenschaft
- Wie macht man Daten F.A.I.R.?
 - FAIRification-Prozess
- Anwendung der F.A.I.R.-Prinzipien auf DL-Modelle für EO-Daten
- Schlussfolgerung
- Quellen

Einleitung

“Data that is loved tends to survive.”

— *Kurt Bollacker*

Die F.A.I.R.-Prinzipien (Findable, Accessible, Interoperable, Reusable) wurden 2016 entwickelt, um den Umgang mit wissenschaftlichen Daten zu verbessern und deren Nutzung in der Forschung zu erleichtern.

In einer Zeit, in der Datenmengen und -komplexität stetig wachsen, sind sie von entscheidender Bedeutung.

FINDABLE

ACCESSIBLE

F.A.I.R.
DATA

INTEROPERABLE

REUSABLE

FINDABLE (Auffindbar)

F1. (Meta-)Daten werden mit einer weltweit eindeutigen und dauerhaften Kennung versehen

F2. Daten werden mit umfangreichen Metadaten beschrieben (definiert durch R1 unten)

F3. Metadaten enthalten eindeutig und explizit den Identifikator der Daten, die sie beschreiben

F4. (Meta-)Daten sind in einer durchsuchbaren Ressource registriert oder indexiert

ACCESSIBLE (Zugänglich)

A1. (Meta-)Daten sind anhand ihrer Kennung über ein standardisiertes Kommunikationsprotokoll abrufbar

A1.1 das Protokoll ist offen, frei und universell implementierbar

A1.2 das Protokoll ermöglicht ein Authentifizierungs- und Autorisierungsverfahren, falls erforderlich

A2. die Metadaten sind zugänglich, auch wenn die Daten nicht mehr verfügbar sind

INTEROPERABLE (Interoperabel)

I1. (Meta-)Daten verwenden eine formale, zugängliche, gemeinsame und breit anwendbare Sprache zur Wissensdarstellung

I2. (Meta-)Daten verwenden Vokabulare, die den F.A.I.R.-Grundsätzen folgen

I3. (Meta-)Daten enthalten qualifizierte Verweise auf andere (Meta-)Daten

REUSABLE (Wiederverwendbar)

R1. (Meta-)Daten sind mit einer Vielzahl von genauen und relevanten Attributen reichhaltig beschrieben

R1.1. (Meta-)Daten werden mit einer klaren und zugänglichen Datennutzungslizenz freigegeben

R1.2. (Meta-)Daten sind mit einer detaillierten Dokumentation über Herkunft, Entstehungsprozess und sämtliche Änderungen verbunden

R1.3. (Meta-)Daten entsprechen den für den Bereich relevanten Gemeinschaftsstandards

Kurzes Anwendungsbispiel

- **Findable**: Klimadaten sind über eindeutige IDs schnell auffindbar.
- **Accessible**: Daten sind öffentlich zugänglich.
- **Interoperable**: Klimadaten können mit anderen Umweltdaten (z.B. Luftqualität) kombiniert werden.
- **Reusable**: Daten sind klar lizenziert und wiederverwendbar für neue Studien.

Unterschied zwischen Open Data und F.A.I.R. Data

- Open Data: Daten sind frei verfügbar
- F.A.I.R. Data: Daten müssen nicht immer öffentlich sein, sollten aber gut dokumentiert und nutzbar gemacht werden.

Warum sollte man Daten F.A.I.R. machen?

- Sicherung von Integrität und Reproduzierbarkeit in der Forschung.
- Förderung der Datenintegration
- Viele Förderer (z. B. UN, WHO, DFG) haben die Prinzipien in ihre Richtlinien aufgenommen.

Bedeutung der F.A.I.R.-Prinzipien allgemein

- Öffentliches Vertrauen
- Ethisches Datenmanagement
- Interoperabilität und Integration
- Nachhaltige Entwicklung
- Anpassungsfähigkeit an neue Technologien

Bedeutung der F.A.I.R.-Prinzipien in der Wissenschaft

- Transparenz und Reproduzierbarkeit
- Förderung der Zusammenarbeit
- Compliance

Wie macht man Daten F.A.I.R.?

Das **Three-Point-Framework** formuliert die wesentlichen Schritte auf dem Weg zu einem globalen Internet aus F.A.I.R.-Daten und -Diensten.

- praktische Anleitungen für Stakeholder
- Framework maximiert die Wiederverwendung vorhandener Ressourcen, erhöht die Interoperabilität und fördert die schnelle Annäherung an Standards und Technologien für F.A.I.R.-Daten und -Dienste

Schritte des FAIRification-Prozesses

- 1. Nicht-FAIR-Daten abrufen**
- 2. Analyse der Daten:** Untersuchung der Inhalte und Struktur der Daten:
Welche Konzepte sind enthalten? Wie sind die Daten strukturiert?
- 3. Semantisches Modell definieren:** beschreibt Bedeutung der
enthaltenen Entitäten und Relationen.
- 4. Daten verlinkbar machen:** Anwendung des semantischen Modells, um
die Daten verlinkbar zu gestalten und so die Interoperabilität und
Wiederverwendbarkeit zu fördern.



Schritte des FAIRification-Prozesses

5. **Lizenz zuweisen.**
6. **Metadaten definieren:** Metadaten unterstützen alle Aspekte von F.A.I.R. und sollten umfassend und aussagekräftig sein.
7. **F.A.I.R.-Ressource bereitstellen:** Die F.A.I.R.-Datenressource (einschließlich Metadaten und Lizenz) wird bereitgestellt, damit die Metadaten durch Suchmaschinen indexiert und die Daten abgerufen werden können.



Anwendung der F.A.I.R.-Prinzipien auf DL-Modelle für EO-Daten

- fördern Open Science und Big Data
 - Nutzung von freier und offener Software (FOSS)
 - öffentliche Evaluationsplattformen
- Bereitstellung von trainierten Modellen und experimentellen Designs in öffentlichen Repositories beschleunigt den Fortschritt, da nicht jeder von Grund auf neu anfangen muss

Ziele der F.A.I.R.-Prinzipien für DL-Modelle in EO

- Systematische Dokumentation und Bereitstellung in durchsuchbaren Repositories
- Einhaltung von Standards für Metadaten, Lizenzen und Dokumentation von Herkunft, Entstehungsprozess und sämtlichen Änderungen
- Nachhaltige Nutzung

Schlussfolgerung

Zusammengefasst stellen die F.A.I.R.-Prinzipien sicher, dass Daten nicht nur für aktuelle Forschungsfragen, sondern auch für zukünftige Generationen von Forschenden verfügbar und verständlich bleiben.

Quellen

- Zitat Bollacker: <https://www.americanscientist.org/article/beautiful-data>
- F.A.I.R. Prinzipien: <https://www.go-fair.org/fair-principles/>
- How to go F.A.I.R.: <https://www.go-fair.org/how-to-go-fair/>
- FAIRification Process: <https://www.go-fair.org/fair-principles/fairification-process/>
- YouTube: How to be F.A.I.R. with your Data: https://youtu.be/5OeCrQE3HhE?si=LC3114uYoc94_M8-
- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., et al. (2016). The FAIR guiding principles for scientific data management and stewardship. *Scientific Data*, 3, 160018. <https://doi.org/10.1038/sdata.2016.18>
- Engelhardt, C. (2022). How to be FAIR with your data. <https://doi.org/10.17875/gup2022-1915>
- Dimitrovski, I., Kitanovski, I., Kocev, D., & Simidjievski, N. (2023). Current trends in deep learning for Earth Observation: An open-source benchmark arena for image classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, 197, 18-35. <https://doi.org/10.1016/j.isprsjprs.2023.01.014>
- Persello, C., Wegner, J. D., Hänsch, R., Tuia, D., Ghamisi, P., & Koeva, M. (2022). Deep learning and Earth observation to support the sustainable development goals: Current approaches, open challenges, and future opportunities. *IEEE Geoscience and Remote Sensing Magazine*, 10(2), 172-200. <https://doi.org/10.1109/MGRS.2021.3136100>
- Yue, P., Shangguan, B., Hu, L., Jiang, L., Zhang, C., Cao, Z., & Pan, Y. (2022). Towards a training data model for artificial intelligence in earth observation. *International Journal of Geographical Information Science*, 36(11), 2113–2137. <https://doi.org/10.1080/13658816.2022.2087223>

Additional Content

- Wie misst man, wie F.A.I.R. Daten sind?
→ FAIR Score

„Mit dem FAIR Score können wir die Nachnutzbarkeit von Forschungsdaten auf institutioneller Ebene bewerten. Unseres Wissens gibt es weltweit keine andere Einrichtung, die die FAIRness der von ihren Forschenden publizierten Forschungsdaten evaluiert und offenlegt.“
— *Freie Universität Berlin*

Schätzfrage: Wie hoch ist der FAIR Index für die Charité im Vergleich zu allgemeinen Repositories?

