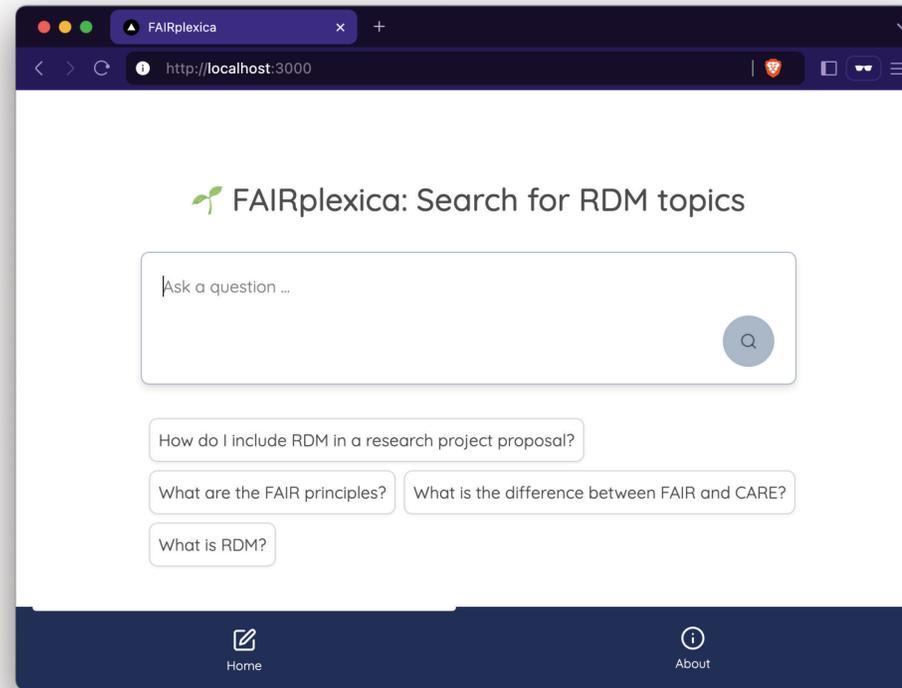
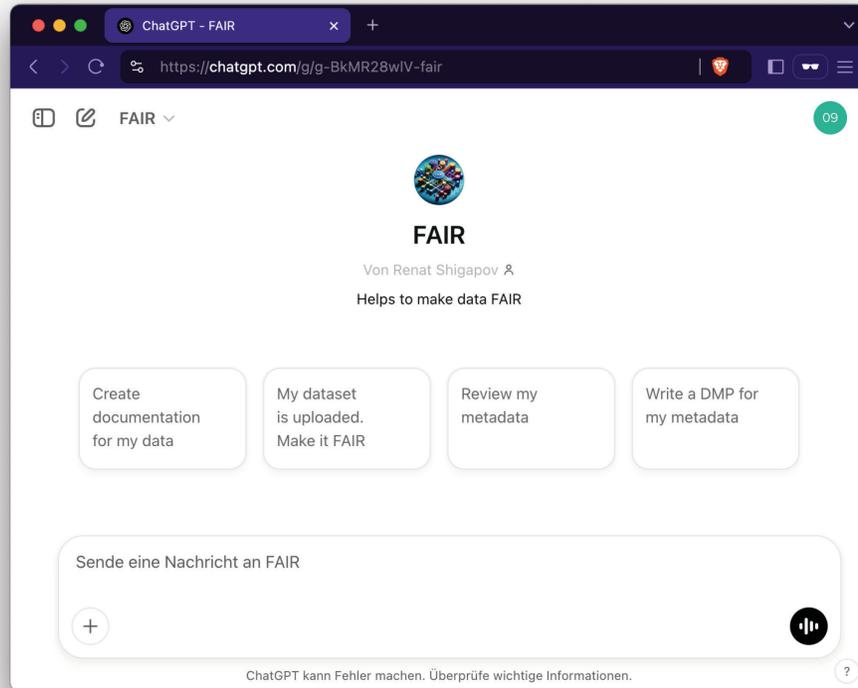


Open-Source-Assistenten für FDM und Open Science: Beratungsservices mit Metasearch und lokalen LLMs



Agenda

- 1. Einführung**
- 2. FAIR GPT**
- 3. FAIRplexica**
- 4. Fazit**

Einführung

Warum virtuelle Assistenten für FDM und Open Science (OS)?

- **24/7-Support:** Fragen zu FDM und OS sofort beantworten
- **Automatisierung:** FDM-Routineaufgaben automatisieren
- **Zugänglichkeit:** verständliche Informationen bereitstellen
- **Aktualität:** aktuelle Informationen bereitstellen

FAIR GPT

FAIR GPT: Was ist FAIR GPT?

- **Seit 2023 verfügbar**
- **Custom GPT in ChatGPT Plus:** Ein virtueller Berater für Forschungsdatenmanagement und Open Science, der externe APIs und integrierte Wissensquellen nutzt, um fortschrittliche Beratung und Lösungen bereitzustellen.

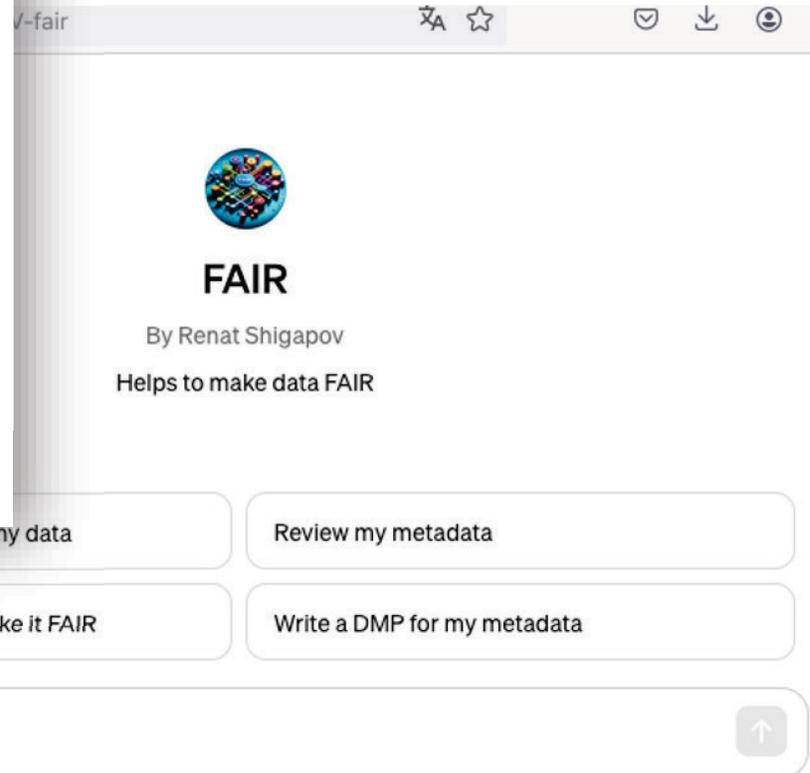


FAIR GPT: <https://chatgpt.com/g/g-BkMR28wIV-fair?model=gpt-4o>

FAIR GPT: Warum FAIR GPT?

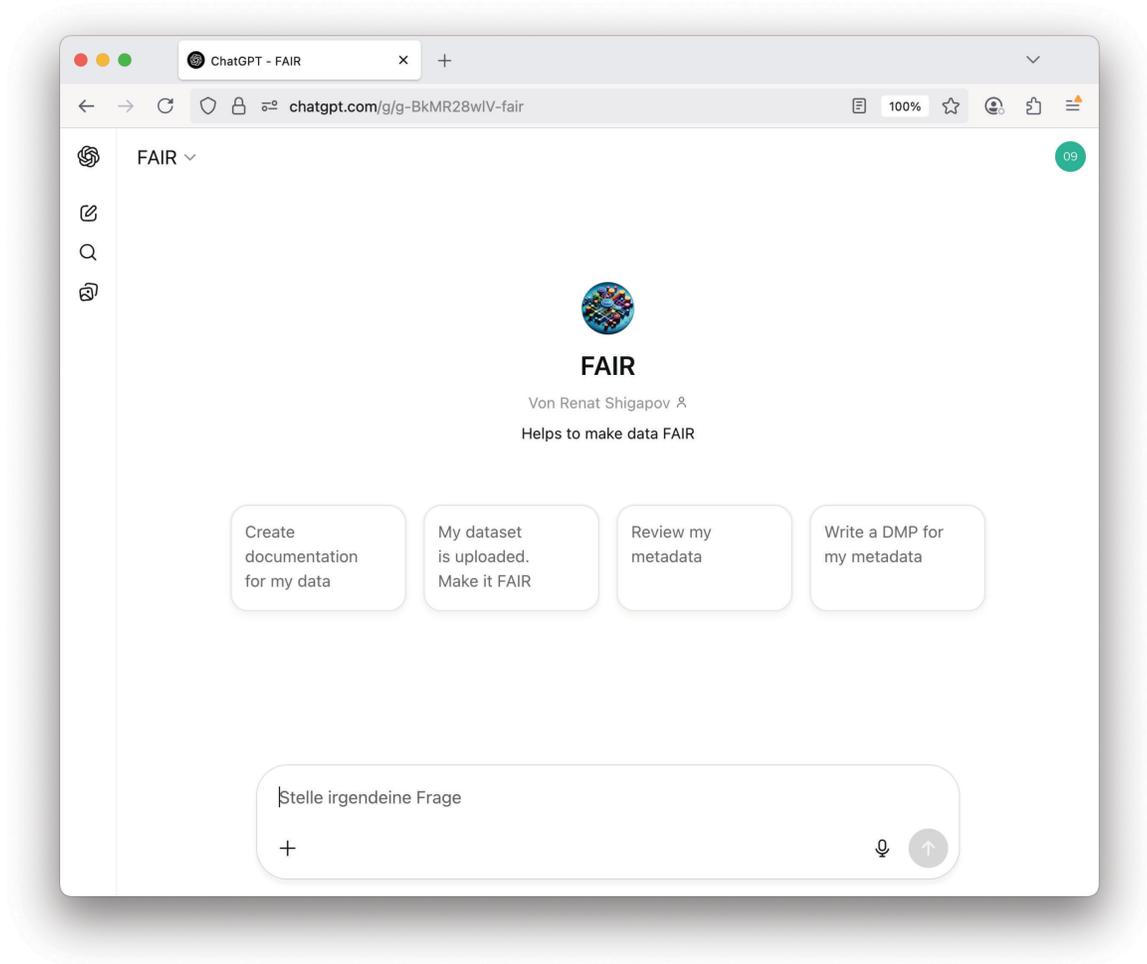


FAIR GPT: <https://chatgpt.com/g-g-BkMR28wIV-fair?model=gpt-4o>



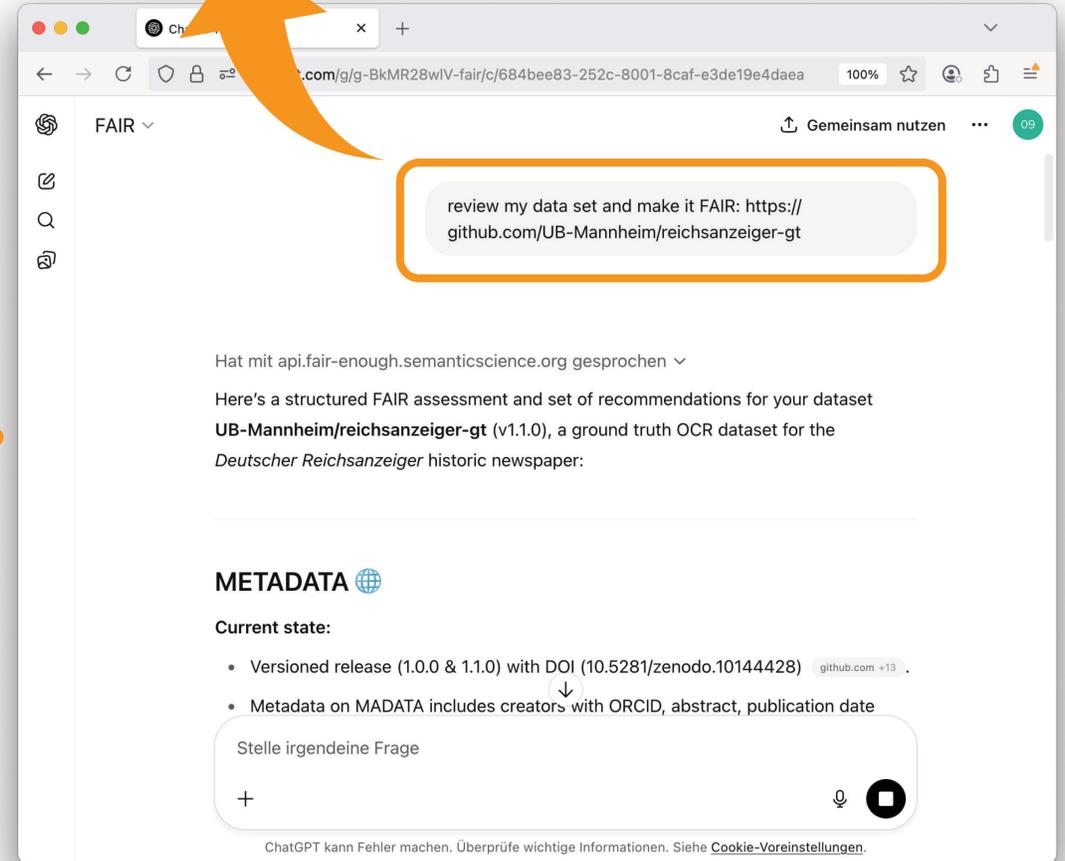
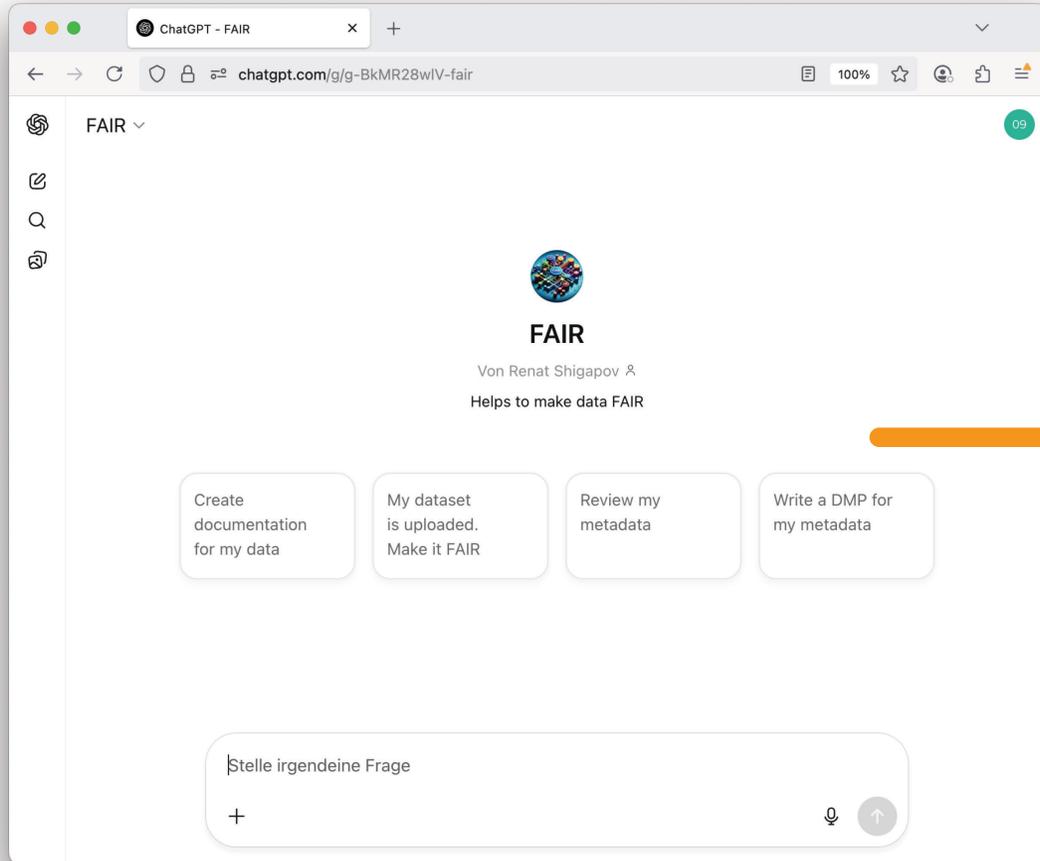
- Metadaten-Review
- Dokumentationshilfe
- Repository-Empfehlungen
- FAIR-Prinzipien-Bewertung
- Datenmanagementplanung
- Datenorganisation
- Lizenzberatung
- Ontologie- und Vokabular-Empfehlungen

FAIR GPT: Demo



FAIR GPT: Demo

review my data set and make it FAIR: <https://github.com/UB-Mannheim/reichsanzeiger-gt>



FAIR GPT: Demo

review my data set and make it FAIR: <https://github.com/UB-Mannheim/reichsanzeiger-gt>



DATA

Current state: PAGE-XML files with layout annotations, text baselines, ground truth lines, and scan URLs via bash-script. [github.com +12](#) [zenodo.org +1](#) [madata.bib.uni-mannheim.de +1](#)

[madoc.bib.uni-mannheim.de](#)

Recommendations:

- Organise in structured directories:

```
swift  
  
/images/ (scans)  
/xml/ (PAGE-XML, one-per-page)  
/lines/ (metadata for each text line)  
/scripts/ (download_images.sh, processing code)
```

- Use consistent filenames (e.g., <YYYYMMDD>_<pageID>.xml).
- Include a comprehensive README and script documentation.
- Add versioned CHANGELOG.
- Add codebooks describing XML schemes, text baselines, special

Stelle irgendeine Frage

ChatGPT kann Fehler machen. Überprüfe wichtige Informationen. Siehe [Cookie-Voreinstellungen](#).

DOCUMENTATION

- README.md exists but could be enriched with sections on:
 - Project overview
 - Data download & setup instructions
 - XML schema reference
 - Transcription guidelines summary
 - Known limitations (only ~197 pages, Fraktur challenges)
 - Licensing info and citation instructions.
- Provide LICENSE, CONTRIBUTING.md.
- Add a codebook (e.g., CODEBOOK.md) listing XML elements and annotation schema.
- Provide example XML excerpt.
- Document data collection and processing pipeline in detail (Transkribus, Kodym & Hradis adaptation).

Stelle irgendeine Frage

ChatGPT kann Fehler machen. Überprüfe wichtige Informationen. Siehe [Cookie-Voreinstellungen](#).

FAIR GPT: Vorteile und Nachteile

Vorteile

- Daten hochladen und Fragen zu den Daten stellen
- Unterstützung beim FDM für Forschende und FDM-Dienste
- Verknüpfung mit externen APIs

Nachteile

- Halluzinationen & falsche Empfehlungen
- Fehlende Transparenz der Herkunft von Empfehlungen
- Datenschutzbedenken
- Keine API-Schnittstelle (=> kann nicht direkt in institutionelle FDM-Workflows integriert werden)



FAIRplexica: Warum FAIRplexica?

→ Nachteile von FAIR GPT verbessern

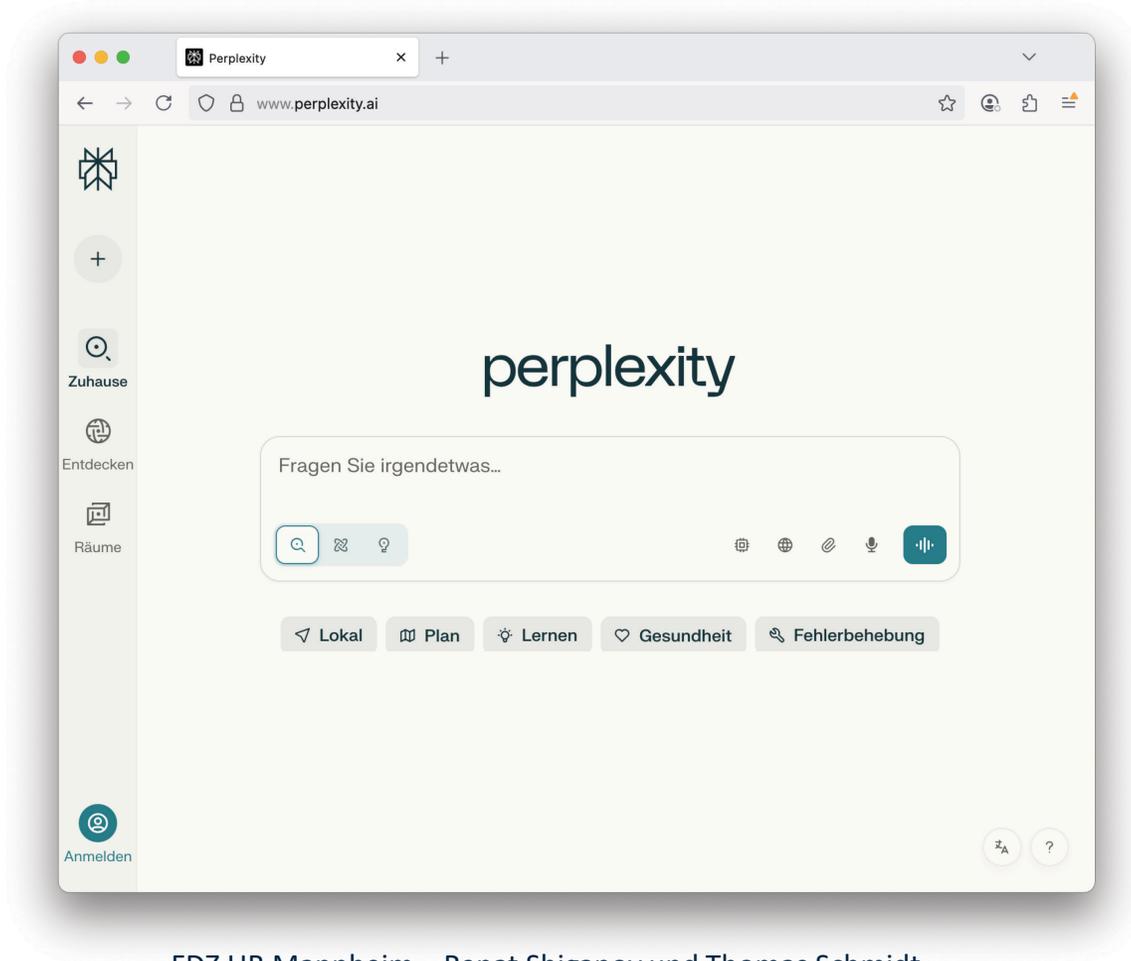
- **Transparenz:** Quellverweise statt opakem LLM-Output
- **Vertrauen:** LLM-Halluzinationen durch Quellen reduzieren
- **Offenheit:** Open Source statt proprietärer, kostenpflichtiger LLM-Service
- **Datenschutz:** Daten der Nutzenden schützen

FAIRplexica: Was ist FAIRplexica?

- **Virtueller Assistent** für Fragen rund um das FDM, **der gezielt Internetressourcen nutzt**, um mit der Unterstützung von **LLMs** hilfreiche Antworten zu formulieren
- Grundidee: **Perplexity.ai** <https://www.perplexity.ai/>



FAIRplexica: Perplexity.ai Demo



FDZ UB Mannheim – Renat Shigapov und Thomas Schmidt
Mittwoch, 25.06.2025

FAIRplexica: P

Perplexity Quellen · 9 Aufgaben

forschungsdaten.info Was ist Forschungsdatenmanag...

Wikimedia Foundati... Forschungsdatenmanage - Wikipedia

TUdata – TU Darmst... Was ist Forschungsdatenmanag...

uni-tuebingen.de Was ist FDM? | Universität Tübingen

perplexity

Fragen Sie irgendetwas...

Lokal Plan Lernen Gesundheit Fehlerbehebung

was ist forschungsdatenmanagement

Perplexity Quellen · 9 Aufgaben

forschungsdaten.info Was ist Forschungsdatenmanag...

Wikimedia Foundati... Forschungsdatenmanage - Wikipedia

TUdata – TU Darmst... Was ist Forschungsdatenmanag...

uni-tuebingen.de Was ist FDM? | Universität Tübingen

Was ist Forschungsdatenmanagement (FDM)?

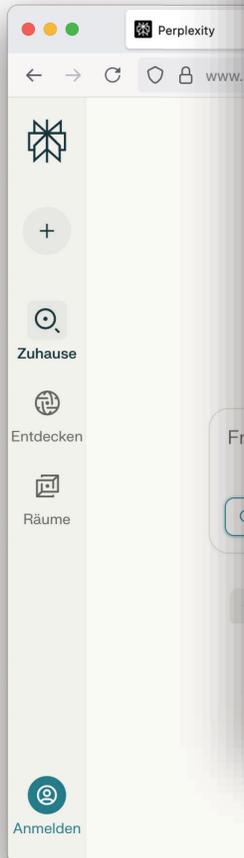
Forschungsdatenmanagement (FDM) bezeichnet alle methodischen, konzeptionellen, organisatorischen und technischen Maßnahmen, die den Umgang mit Forschungsdaten über deren gesamten Lebenszyklus hinweg regeln ^{2 3 4}. Ziel ist es, Forschungsdaten so zu organisieren, dass sie langfristig zugänglich, nachvollziehbar, nachnutzbar und überprüfbar bleiben ^{1 5 6}.

Zentrale Aspekte des Forschungsdatenmanagements

- **Planung und Generierung:** Bereits bei der Planung eines Forschungsprojekts beginnt FDM, etwa durch die Erstellung eines Datenmanagementplans, der den Umgang mit den entstehenden Daten regelt ^{1 3 6}.
- **Organisation und Dokumentation:** Forschungsdaten werden strukturiert erfasst,

Fragen Sie irgendetwas...

FAIRplexica: P



Perplexity Quellen · 9 Aufgaben

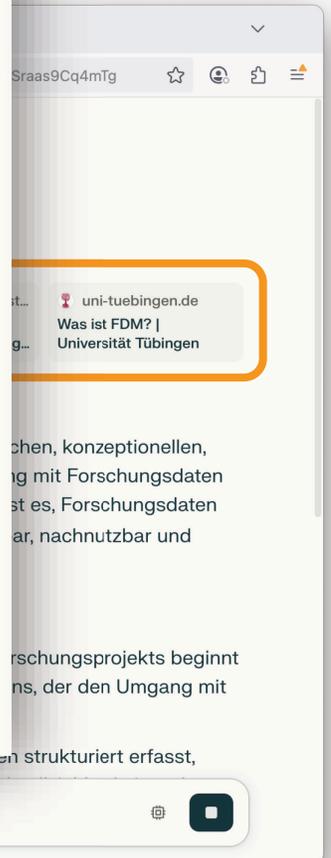
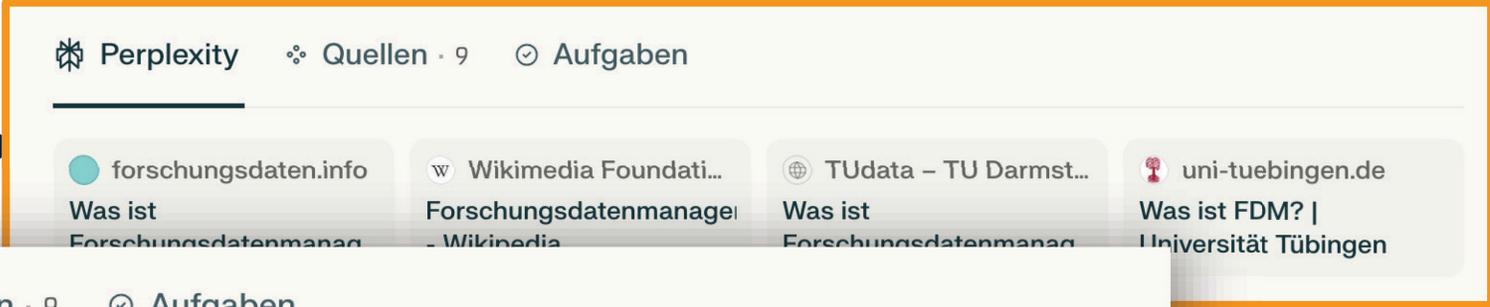
forschungsdaten.info Was ist Forschungsdatenmanag...
Wikimedia Foundati... Forschungsdatenmanage... - Wikipedia
TUdata – TU Darmst... Was ist Forschungsdatenmanag...
uni-tuebingen.de Was ist FDM? | Universität Tübingen

Was ist Forschungsdatenmanagement (FDM)?

Forschungsdatenmanagement (FDM) bezeichnet alle methodischen, konzeptionellen, organisatorischen und technischen Maßnahmen, die den Umgang mit Forschungsdaten über deren gesamten Lebenszyklus hinweg regeln ² ³ ⁴. Ziel ist es, Forschungsdaten so zu organisieren, dass sie langfristig ¹ ⁵ ⁶ überprüfbar bleiben.

Zentrale Aspekte des Forschungsdatenmanagements

uni-tuebingen.de Was ist FDM? | Universität Tübingen



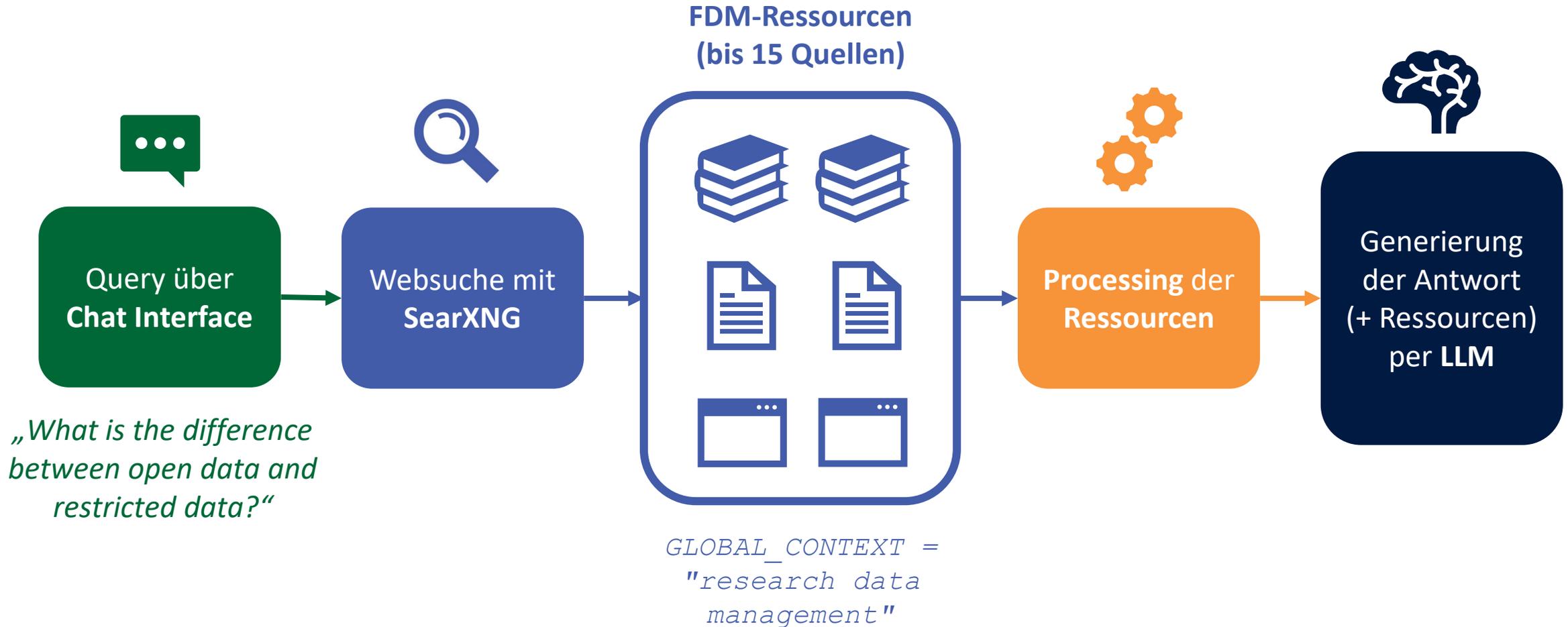
FAIRplexica: Was ist FAIRplexica?

- **Perplexity.ai:** proprietäre Web-App (kostenlos + „Pro“-Version)
- **Idee:** auf GitHub verfügbare Open-Source-Varianten von Perplexity.ai für FDM und Open Science anpassen
- 3 dieser Varianten von uns getestet:
 - **FAIR-sensei:** <https://github.com/UB-Mannheim/FAIR-sensei>
 - Fork von <https://github.com/jjleng/sensei>
 - **FAIR-farfalle:** <https://github.com/UB-Mannheim/FAIR-farfalle>
 - Fork von <https://github.com/rashadphz/farfalle>
 - **FAIRplexica:** <https://github.com/UB-Mannheim/FAIRplexica>
 - Fork von <https://github.com/ltzCrazyKns/Perplexica>

FAIRplexica: Vergleich

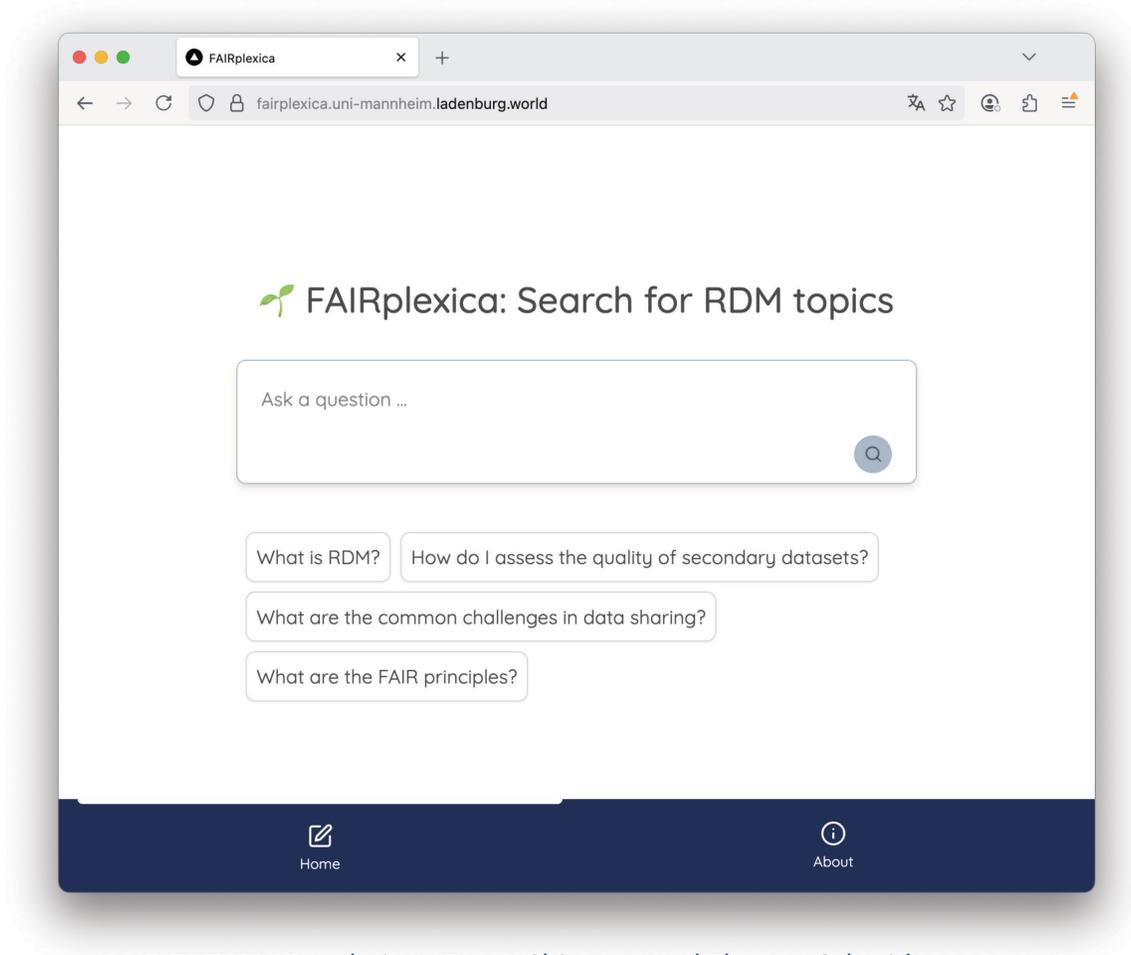
Legende: **positiv**, **neutral**, **negativ**

Kategorien	Sensei	Farfalle	Perplexica
LLM Cloud Services	OpenAI, Anthropic	OpenAI, Groq	OpenAI, Groq, Google, Anthropic
Lokale LLMs	Ja (aber: hard-coded)	Ja (Ollama, litellm)	Ja (Ollama, custom OpenAI API endpoints)
Suchmaschine	SearXNG	SearXNG	SearXNG
Tech-Stack	Next.js, Fast API	Next.js, Fast API	Next.js, Express.js, SQLite
GitHub Stars	446	ca. 3.300	ca. 22.300
Letzter Commit	Okt. 2024	Sep. 2024	Juni 2025
Production-ready	nein	nein	ja
Lizenz	Apache 2.0	Apache 2.0	MIT
Fazit	😊 😊 🤔 🤔 😞 😞 😞	😊 😊 😊 🤔 😞 😞 😞	😊 😊 😊 😊 😊 😊 😊 😞

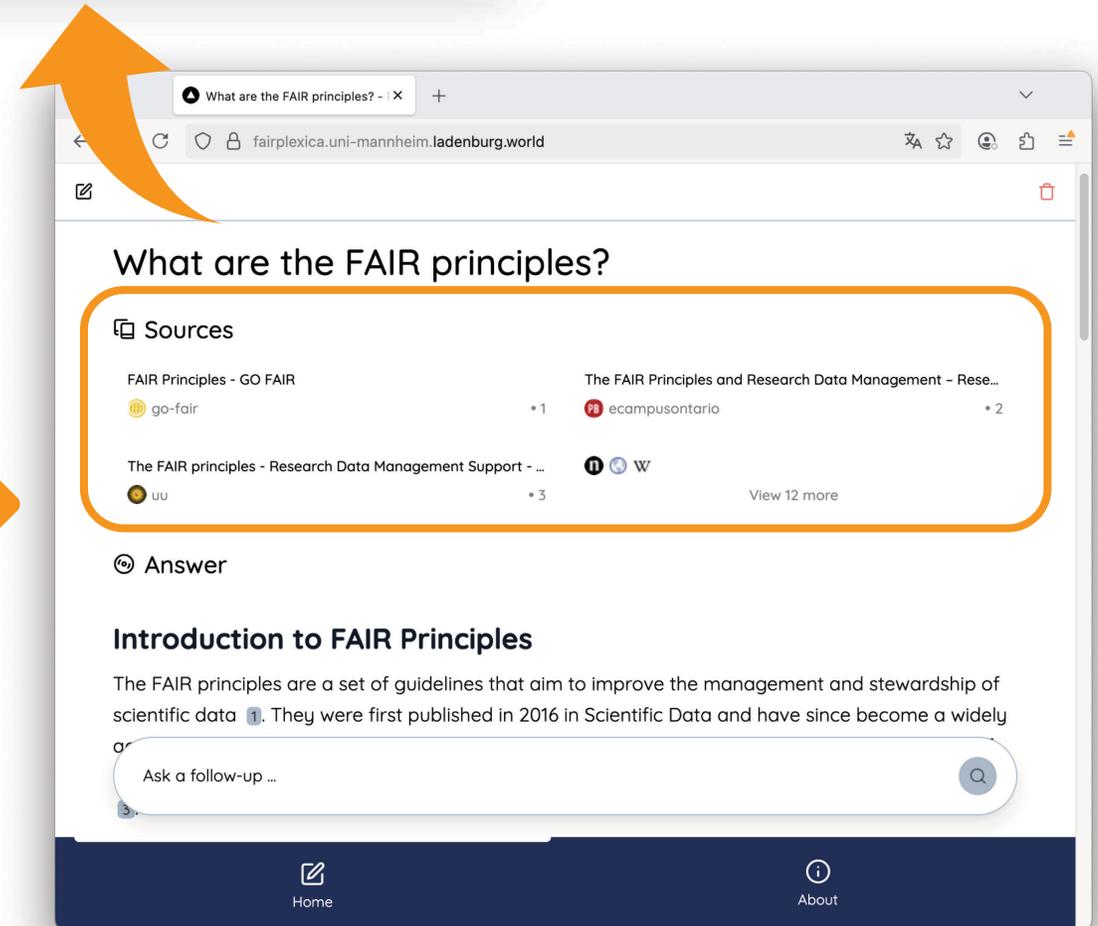
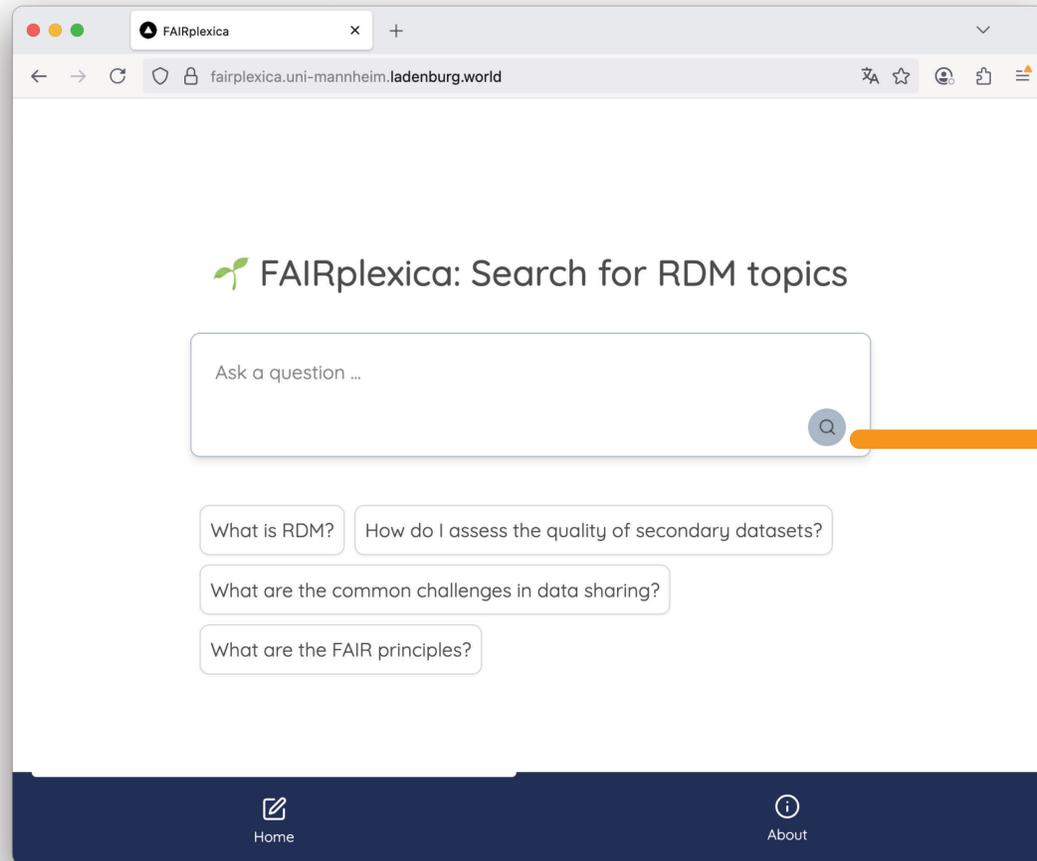


„What is the difference between open data and restricted data?“

FAIRplexica: Demo



FAIRplexica: What are the FAIR principles?



Sources

- | | |
|--|---|
| FAIR Principles - GO FAIR
go-fair • 1 | The FAIR Principles and Research...
ecampusontario • 2 |
| The FAIR principles - Research D...
uu • 3 | The FAIR Guiding Principles for s...
nature • 4 |
| FAIR data - Research Data Man...
ufs • 5 | FAIR data - Wikipedia
en • 6 |
| The FAIR guiding principles for r...
publisso • 7 | How to make your data FAIR
openaire • 8 |
| What are the FAIR Data Principl...
asklib • 9 | FAIR principles for Research - D...
guides • 10 |

What are the FAIR principles?

Sources

- FAIR Principles - GO FAIR (go-fair) • 1
- The FAIR Principles and Research Data Management - Research Data Management Support - ... (ecampusontario) • 2
- The FAIR principles - Research Data Management Support - ... (uu) • 3

View 12 more

Answer

Introduction to FAIR Principles

The FAIR principles are a set of guidelines that aim to improve the management and stewardship of scientific data. They were first published in 2016 in Scientific Data and have since become a widely...

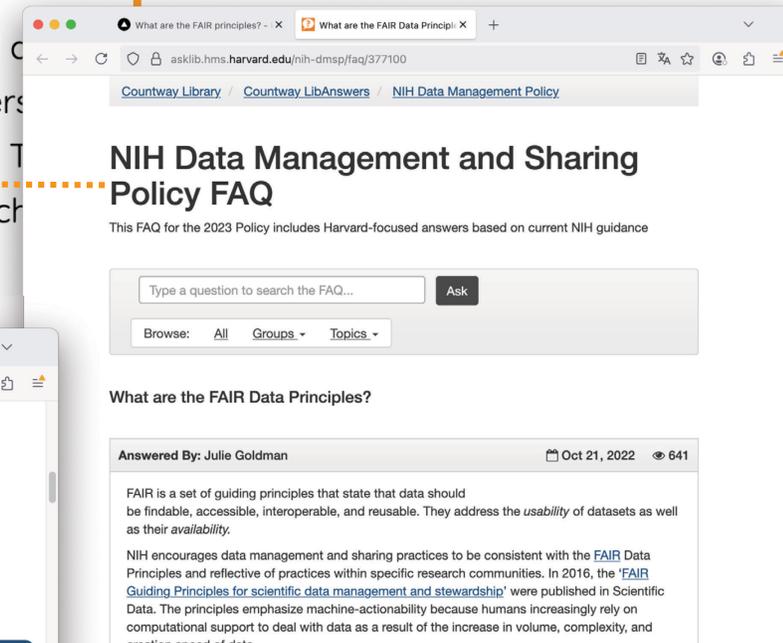
Ask a follow-up ...

Importance of FAIR Principles

The FAIR principles are important because they help to ensure that research data is managed and shared in a way that is efficient, effective, and sustainable ⁸. By making data FAIR, researchers can increase the impact and value of their research, and help to accelerate scientific discovery ⁹. The FAIR principles also help to promote transparency, reproducibility, and collaboration in research, which are essential for advancing knowledge and addressing complex challenges ¹⁰.

Importance of FAIR Principles

The FAIR principles are important because they help to ensure that research data is managed and shared in a way that is efficient, effective, and sustainable ⁸. By making data FAIR, researchers increase the impact and value of their research, and help to accelerate scientific discovery ⁹. The FAIR principles also help to promote transparency, reproducibility, and collaboration in research, which are essential for advancing knowledge and addressing complex challenges ¹⁰.



Countway Library / Countway LibAnswers / NIH Data Management Policy

NIH Data Management and Sharing Policy FAQ

This FAQ for the 2023 Policy includes Harvard-focused answers based on current NIH guidance

Type a question to search the FAQ...

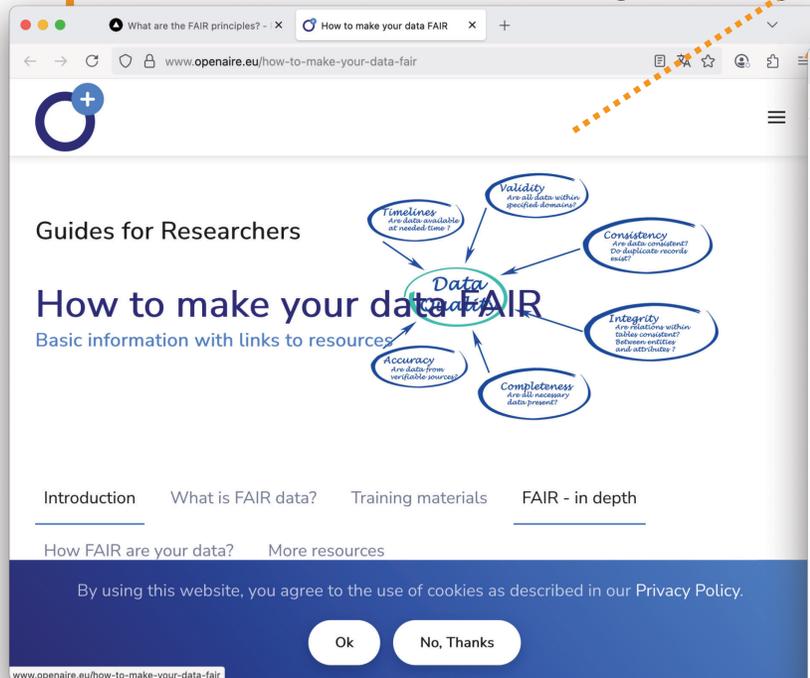
Browse: All Groups Topics

What are the FAIR Data Principles?

Answered By: Julie Goldman Oct 21, 2022 641

FAIR is a set of guiding principles that state that data should be findable, accessible, interoperable, and reusable. They address the *usability* of datasets as well as their *availability*.

NIH encourages data management and sharing practices to be consistent with the FAIR Data Principles and reflective of practices within specific research communities. In 2016, the 'FAIR Guiding Principles for scientific data management and stewardship' were published in Scientific Data. The principles emphasize machine-actionability because humans increasingly rely on computational support to deal with data as a result of the increase in volume, complexity, and creation speed of data.



Guides for Researchers

How to make your data FAIR

Basic information with links to resources

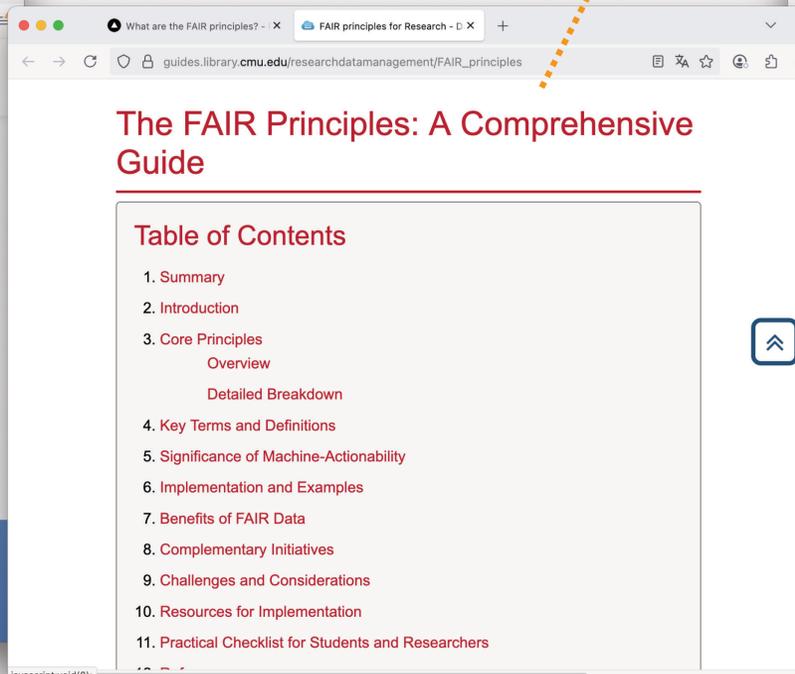


Introduction | What is FAIR data? | Training materials | FAIR - in depth

How FAIR are your data? | More resources

By using this website, you agree to the use of cookies as described in our Privacy Policy.

www.openaire.eu/how-to-make-your-data-fair



The FAIR Principles: A Comprehensive Guide

Table of Contents

1. Summary
2. Introduction
3. Core Principles
 - Overview
 - Detailed Breakdown
4. Key Terms and Definitions
5. Significance of Machine-Actionability
6. Implementation and Examples
7. Benefits of FAIR Data
8. Complementary Initiatives
9. Challenges and Considerations
10. Resources for Implementation
11. Practical Checklist for Students and Researchers

javascript:void(0);

 **FAIRplexica – Vorteile**

- Alle Komponenten von FAIRplexica sind **Open Source** (MIT Lizenz):
 - **Websuche** mit *SearXNG*
 - **Ressourcen-Processing** und **Generierung von Antworten** mit ...
 - lokalen LLMs (*Ollama*) oder
 - anderen, OpenAI analogen, APIs (bspw. *GWDG*)
 - **Backend**
 - **Frontend** mit GUI

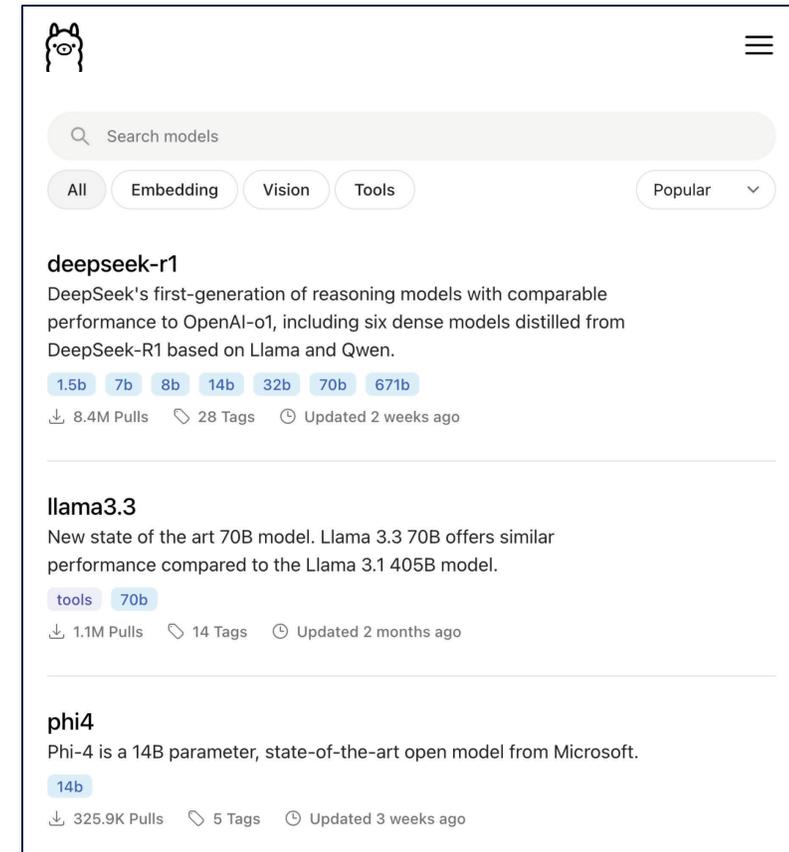


- **Metasuch-Engine**, die Ergebnisse verschiedener Suchmaschinen aggregiert
 - SearXNG sendet Suchanfragen anonym an verschiedene Suchmaschinen (bspw. Google, Bing, etc.) und kombiniert deren Ergebnisse
 - Keine Weitergabe von IP-Adressen
 - Keine Erstellung von Profilen der Nutzenden
 - <https://docs.searxng.org/>

FAIRplexica – Vorteile: lokale LLMs



- Open-Source-LLMs können mit **Ollama** verwendet werden
 - Zahlreiche LLMs verfügbar
 - <https://github.com/ollama/ollama>



FAIRplexica – weitere Vorteile

- **Datenschutz:** Bei Verwendung lokaler LLMs werden Fragen von Nutzenden **nicht** über Provider (bspw. OpenAI) geleitet
- **Sprachagnostisch:** Fragen auf Deutsch, Englisch, Französisch etc.
- Inhaltliche **Vertiefungsebene** durch „related“ Questions



FAIRplexica: <https://github.com/UB-Mannheim/FAIRplexica>

FAIRplexica – Einschränkungen

FAIRplexica – Einschränkungen

- **Performance** hängt von Hardware ab (bei Verwendung lokaler LLMs)
- **Qualität** der Antworten ist von Größe des LLM abhängig
- LLMs können sich irren, bzw. Ressourcen **nicht vollständig** oder **falsch verstehen**



FAIRplexica: <https://github.com/UB-Mannheim/FAIRplexica>

Fazit

Fazit

- Virtuelle Assistenten können **Beratungsservices** im FDM und Open-Science-Bereich unterstützen
- FAIRplexica verbessert **Transparenz** der Antworten durch Verweis auf Internetressourcen
- **Besserer Datenschutz** bei Verwendung lokaler LLMs
- FAIRplexica kann auf **eigener Infrastruktur** lokal betrieben werden

Vielen Dank für Ihre Aufmerksamkeit!

Renat Shigapov : renat.shigapov@uni-mannheim.de

Thomas Schmidt: thomas.schmidt@uni-mannheim.de

Quellen und Ressourcen

1. Shigapov, R., & Schumm, I. (2024). FAIR GPT: A virtual consultant for research data management in ChatGPT. arXiv. <https://doi.org/10.48550/arXiv.2410.07108>
2. FAIR GPT at GitHub: <https://github.com/UB-Mannheim/FAIR-GPT>
3. FAIR GPT at OpenAI: <https://chat.openai.com/g/g-BkMR28wIV-fair>
4. FAIRplexica at GitHub: <https://github.com/UB-Mannheim/FAIRplexica/tree/config-api-server>
5. SearXNG: <https://docs.searxng.org/>
6. FAIR-farfalle at GitHub: <https://github.com/UB-Mannheim/FAIR-farfalle>
7. FAIR-Perplexica at GitHub: <https://github.com/UB-Mannheim/FAIR-Perplexica>
8. FAIR-sensei at GitHub: <https://github.com/UB-Mannheim/FAIR-sensei>