



**PROFESSUR  
MEDIENINFORMATIK**

**Literatur & Daten**

Methodenseminar Informatik für Geistes- und Sozialwissenschaftler  
Maximilian Eibl, Medieninformatik, TU Chemnitz



## Wozu Literatur?

Die Berücksichtigung wissenschaftlicher Literatur hilft Ihnen:

- Eine Idee zu entwickeln, die Ihrer Arbeit zugrunde liegt,
- die wissenschaftliche Relevanz ihrer Frage zu beurteilen und belegen,
- Ihre eigene Arbeit klar von Vorarbeiten abzugrenzen,
- Ihren Beitrag für die Wissenschaft klar herauszustellen und
- alternative Ansätze begründet zu verwerfen.

„Bernhard von Chartres sagte, wir seien gleichsam **Zwerge**, die **auf den Schultern von Riesen** sitzen, um mehr und Entfernteres als diese sehen zu können – freilich nicht dank eigener scharfer Sehkraft oder Körpergröße, sondern weil die Größe der Riesen uns emporhebt.“

– Johannes von Salisbury (1159). *Metalogicon* 3,4,47–50



## Arten von Literatur

- *Fach- und Lehrbücher*
- *Wissenschaftliche Zeitschriften (engl.: scientific journals)*
- *Tagungsbände (engl.: proceedings)*
- *Graue Literatur*
  
- *Webseiten / Wikipedia*



## Beurteilung einer Quelle

- *Begutachtung*
- *Organisation*
- *Zitierung*
- *Autoren*



# Bibliothek

- [FAQ](#) zur Bibliothek
- [Foliensätze](#) mit den wichtigsten Informationen rund um die Bibliothek
- [BIB4U](#) - Hier finden Sie zahlreiche Infos zur Nutzung der Bibliothek und ihrer Services.
- [IKOnline](#) - der Selbstlernkurs Informationskompetenz vertieft die Informationen zur Arbeit mit Literatur
- [Tutorials](#) - kurze Videos zu speziellen Angeboten



- [Online Public Access Catalogue \(OPAC\)](#)
- [Neuer Katalog](#)
- [Regionalkatalog Sachsen](#)
- [Fernleihe](#)
- [Elektronische Zeitschriftenbibliothek \(EZB\)](#)
- [Datenbank-Informationssystem \(DBIS\)](#)



## Vorsicht Plagiat

Grundsätzlich ist die Verwendung von fremden Ideen und Zitaten in der Wissenschaft notwendig, nur darf man sie sich nicht aneignen: Unter Plagiat versteht man die Übernahme von fremdem geistigen Eigentum, ohne diese Übernahme durch Referenzierung (s.o.) deutlich zu machen. Dazu gehören im einfachsten (und häufigsten) Fall **wörtlich abgeschriebene** Textstellen aber auch übernommene Bilder, Diagramme, im Netz auch Videos und Tonaufnahmen. Plagiate können aber auch **Paraphrasierungen oder Übersetzungen** sein, wenn deren Ursprung nicht deutlich gemacht wird. Rein rechtlich gesehen handelt es sich dabei um einen Verstoß gegen das Urheberrechtsgesetz und damit um Diebstahl. Im Wissenschaftsbetrieb und damit auch im Studium ist ein Plagiat ein sehr ernstes Vergehen und ein K.O.-Kriterium für die Arbeit, in der es vorkommt, d.h. **eine Arbeit, egal ob Proseminararbeit oder Dissertation, wird automatisch mit der Note 5 bewertet, wenn plagiiert wurde.** Einfache Plagiate sind durch automatische Plagiatsprüfer schnell zu identifizieren. Wenn Sie mehr über Plagiate und Plagiatsjäger wissen wollen, besuchen Sie [VroniPlag](#).



# Forschungsdatenmanagement



English | Direktlink | Mein Profil | Kontakt | Suchwort | Suche in... | Q

TECHNISCHE UNIVERSITÄT  
IN DER KULTLANDSCHAFT ERZBERG  
CHEMNITZ

Universität | Fakultäten | Zentrale Einrichtungen | Studium | International

Universitätsbibliothek

TU Chemnitz → Universitätsbibliothek → Publizieren → Forschungsdatenmanagement

Online-Katalog | SUCHWORT

Publizieren

## Forschungsdatenmanagement

von der Planung bis zur Nachnutzung von Forschungsdaten

Forschungsdatenmanagement (FDM) umfasst alle Maßnahmen im Lebenszyklus von Forschungsdaten: von der Planung, Erhebung, Speicherung, Dokumentation, Aufbereitung und Analyse der Daten bis zu ihrer Veröffentlichung, Archivierung und Nachnutzung.

Ein verantwortungsvoller, transparenter und strukturierter Umgang mit Forschungsdaten ist nicht nur Teil der guten wissenschaftlichen Praxis sondern ermöglicht auch eine langfristige Nachnutzung und Reproduzierbarkeit von wissenschaftlichen Ergebnissen.

Neben der Organisation, Dokumentation und Archivierung von Forschungsdaten soll FDM auch zur öffentlichen Bereitstellung und langfristigen Nachnutzung von Daten beitragen, weshalb im Zusammenhang mit FDM oft die Themen FAIR und Open Data genannt werden.

<b>Forschungsdaten</b> 	<b>Datenlebenszyklus</b> 	<b>Datenmanagementplan</b> 
<b>Datenverwaltung</b> 	<b>Datenveröffentlichung</b> 	<b>Rechtliche Aspekte</b> 

→ Fragen | Anregungen | Kritik  
→ Wörterbuch Deutsch - Englisch  
→ Bildnachweise  
→ Mastodon  
→ Twitter/X  
→ Instagram  
→ ubLOG

**Sie haben Fragen rund ums FDM?**  
Wir beantworten diese gern in einem individuellen Gespräch.  
Buchten Sie einen Termin über → [Book a Librarian](#)  
Oder schreiben Sie direkt eine Nachricht an Ihre Ansprechpartnerin:  
Frau → [Anja Hähle](#)

**Informationen**  
 Allgemein  
 Materialien

**News**  
 Aktuell informiert  
 Anders informiert

OPEN LIBRARY  
BADGE 2020

Notfall | Kontakt | Campusplan | Universität | Chemnitz

© 2024 Technische Universität Chemnitz  
Webteam-UB | 6. September 2024

Für Autoren | Impressum | Datenschutz | Zusätzliche Datenschutzerklärung der UB Chemnitz | Cookies  
Barrierefreiheit | Metadaten

<https://www.tu-chemnitz.de/ub/publizieren/forschungsdaten/>

# Forschungsdaten im Datenzyklus

## Vorhaben planen

- Datenmanagement planen (→ [DMP](#) erstellen)
- → [Recherche](#) nach bereits vorhanden Daten

## Daten erheben

- Experimente, Befragungen, Simulationen, usw. durchführen
- Erlaubnis zur Datennutzung einholen

## Daten verarbeiten

- Daten erfassen, digitalisieren, transkribieren, übersetzen
- Daten prüfen, validieren, bereinigen, → [anonymisieren](#)
- → [Daten dokumentieren](#) und mit Metadaten versehen
- → [Daten verwalten](#) und → [sichern](#)
- Daten für die → [Langzeitarchivierung](#) vorbereiten (Dateiformate)
- Daten interpretieren und Schlussfolgerungen ableiten

## Daten veröffentlichen

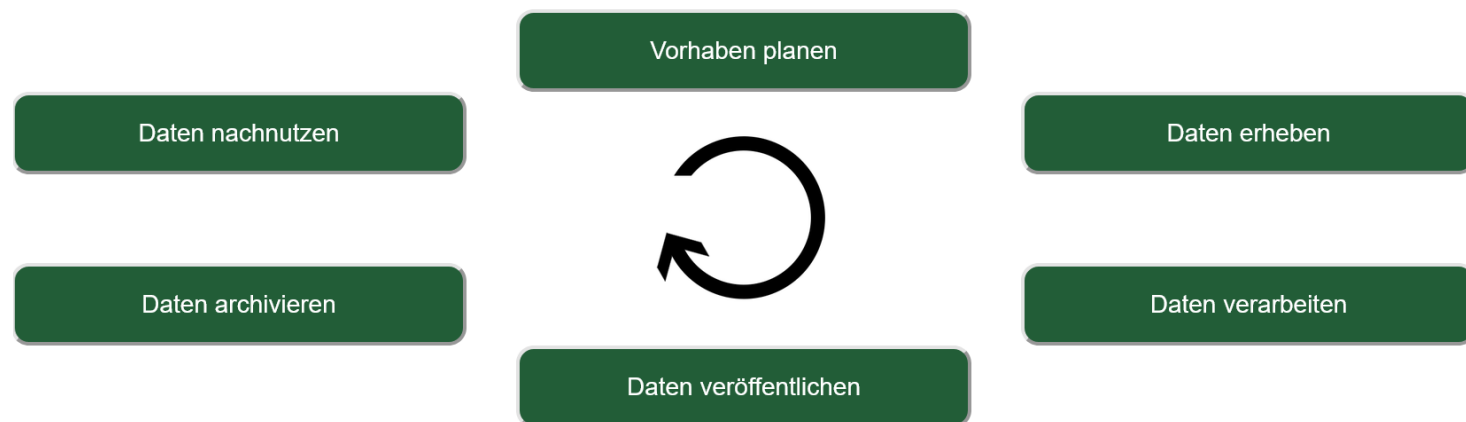
- → [rechtliche und ethische Aspekte](#) klären
- → [Lizenzen](#) festlegen
- Veröffentlichungsort (→ [Repositorium](#), → [Journal](#)) auswählen

## Daten archivieren

- endgültige → [Datenversionen](#) auswählen
- Ort für Langzeitarchivierung auswählen (Datenträger oder offenes Repositorium)

## Daten nachnutzen

- Daten bzw. Schlussfolgerungen prüfen
- Daten für Lehre oder Forschungsvorhaben nutzen



# FAIR Guiding Principles for scientific data management and stewardship

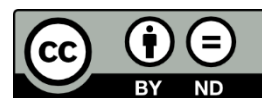
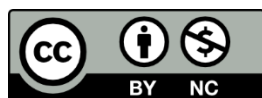
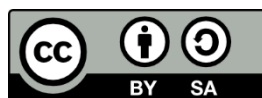


- **Findable**
  - [F1. \(Meta\)data are assigned a globally unique and persistent identifier](#)
  - [F2. Data are described with rich metadata \(defined by R1 below\)](#)
  - [F3. Metadata clearly and explicitly include the identifier of the data they describe](#)
  - [F4. \(Meta\)data are registered or indexed in a searchable resource](#)
- **Accessible**
  - [A1. \(Meta\)data are retrievable by their identifier using a standardised communications protocol](#)
  - [A1.1 The protocol is open, free, and universally implementable](#)
  - [A1.2 The protocol allows for an authentication and authorisation procedure, where necessary](#)
  - [A2. Metadata are accessible, even when the data are no longer available](#)
- **Interoperable**
  - [I1. \(Meta\)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.](#)
  - [I2. \(Meta\)data use vocabularies that follow FAIR principles](#)
  - [I3. \(Meta\)data include qualified references to other \(meta\)data](#)
- **Reusable**
  - [R1. \(Meta\)data are richly described with a plurality of accurate and relevant attributes](#)
  - [R1.1. \(Meta\)data are released with a clear and accessible data usage license](#)
  - [R1.2. \(Meta\)data are associated with detailed provenance](#)
  - [R1.3. \(Meta\)data meet domain-relevant community standards](#)

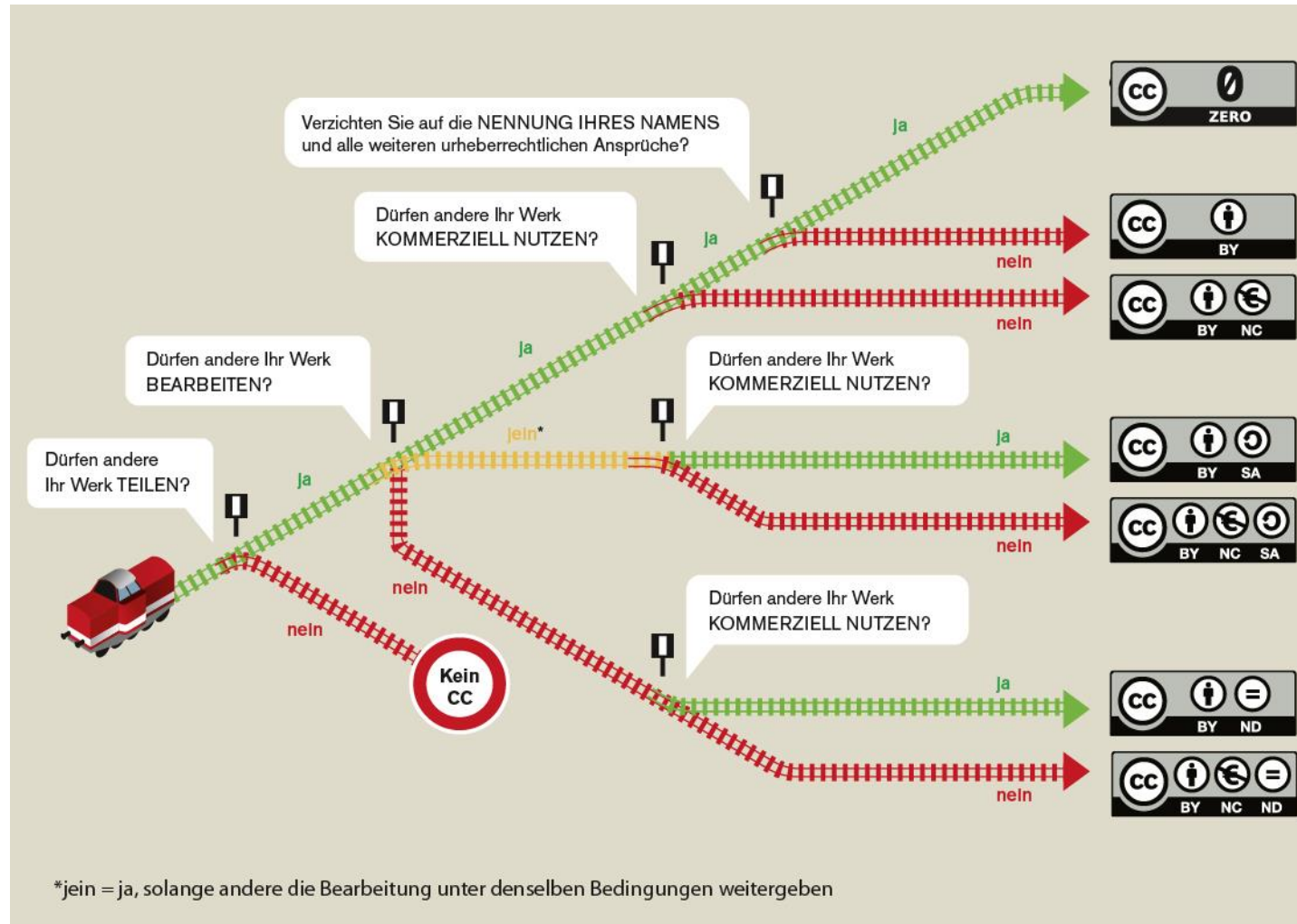


# Lizenzen: Creative Commons

- Ⓘ BY: credit must be given to the creator.
- Ⓒ SA: Adaptations must be shared under the same terms.
- Ⓝ NC: Only noncommercial uses of the work are permitted.
- Ⓔ ND: No derivatives or adaptations of the work are permitted.



# CC Lizenzmodelle



# Weitere Lizenzen



## Institut für Rechtsfragen der Freien und Open Source Software

### ifrOSS Lizenzcenter

#### Kategorien

Im Lizenzcenter des ifrOSS werden die Lizenzen für freie Inhalte gesammelt und verschiedenen Lizenztypen zugeordnet. Dies betrifft folgende Kategorien:

- [Freie Software und Open Source Lizenzen](#)
- [Open Content Lizenzen](#)
- [Open Data Lizenzen](#)
- [Open Hardware Lizenzen](#)
- [Sonstige Freie Lizenzen](#)
- [Ähnliche Lizenzen, die nicht den Anforderungen der Open Source Definition genügen](#)
- [Ethical Lizenzen](#)
- [Public Domain Erklärungen](#)
- [Open AI Lizenzen](#)

Damit wird eine einfache Auffindbarkeit und Charakterisierung ermöglicht. Da fortlaufend neue Lizenzen erstellt werden, in neuen Versionen erscheinen oder nicht mehr unter ihrer alten URL aufrufbar sind, ist die Beteiligung aller bei der Aktualisierung dieser Liste willkommen.

[https://ifross.github.io/ifrOSS/Pages/licence\\_center/de](https://ifross.github.io/ifrOSS/Pages/licence_center/de)

Dieses Projekt wird von den [Mitgliedern](#) des ifrOSS betreut.

[View on GitHub](#)



[ifross.org](#) [Impressum](#) [Datenschutz](#) [Lizenzen](#)

Published January 26, 2024 | Version 0.0.3

Software

## txt-utils

Taubert, Stefan

Show affiliations

For a project description please look at [GitHub/PyPI](#).

### Files

Name	Size	Download all
txt-utils-0.0.3.tar.gz <small>10571273@zenodo.org</small>	50.6 kB	
txt-utils-0.0.3-py3-none-any.whl <small>10571273@zenodo.org</small>	23.3 kB	

### Additional details

**Related works**  
**Is derived from**  
Software: <https://github.com/stefantaubert/txt-utils> (URL)  
**Is identical to**  
Software: <https://pypi.org/project/txt-utils> (URL)

This site uses cookies. Find out more on how we use cookies  
**Funding**  
   
Deutsche Forschungsgemeinschaft

**Citations**

**Show only:**  Literature (0)  Dataset (0)  Software (0)  Unknown (0)  Citations To This Version

No citations found

87 VIEWS    24 DOWNLOADS

[Show more details](#)

### Versions

Version 0.0.3 <small>10.5281/zenodo.10571273</small>	Jan 26, 2024
Version 0.0.2 <small>10.5281/zenodo.7986310</small>	May 30, 2023
Version 0.0.1 <small>10.5281/zenodo.7037645</small>	May 30, 2022

[View all 3 versions](#)

Cite all versions? You can cite all versions by using the DOI 10.5281/zenodo.7037644. This DOI represents all versions, and will always resolve to the latest one. [Read more.](#)

### External resources

Indexed in

OpenAIRE

### Keywords and subjects

[Preprocessing](#) [Processing](#) [Text-to-speech](#)  
[Speech synthesis](#) [Utils](#) [Language](#) [Linguistics](#)

### Details

**DOI**  
[10.5281/zenodo.10571273](https://doi.org/10.5281/zenodo.10571273)

**Resource type**  
Software

**Publisher**  
Zenodo

**Languages**  
English

### Rights

MIT License

### Citation

Taubert, S. (2024). txt-utils (0.0.3). Zenodo. <https://doi.org/10.5281/zenodo.10571273>

Style:

### Export

### Technical metadata

Created January 26, 2024  
Modified January 26, 2024

Jump up



<https://zenodo.org/records/10571273>

# Github

<https://github.com/stefantaubert/txt-utils>

The screenshot displays the GitHub interface for the repository `stefantaubert/txt-utils`. The top navigation bar includes a search bar, a 'Product' menu (circled in red), and links for 'Sign in' and 'Sign up'. The repository page shows the following details:

- Repository Overview:** 146296 files, 10 months ago, 48 commits.
- File List:** A table of files including `src`, `.gitignore`, `CHANGELOG.md`, `CITATION.cff`, `CODE_OF_CONDUCT.md`, `CONTRIBUTING.md`, `LICENSE`, `MANIFEST.in`, `Pipfile`, `Pipfile.lock`, `README.md`, `pylintrc`, `pyproject.toml`, and `setup.cfg`.
- README:** The main content area showing the repository's purpose and features.

**README Content:**

### txt-utils

CLI to batch process lines of a single text file.

CLI to modify text files.

#### Features

- `merge`: merge multiple text files into one
- `extract-vocabulary`: extract unit vocabulary
- `transcribe`: transcribe units
- `replace`: replace text
- `replace-line`: replace text in a line
- `trim-units`: trim units
- `remove-units`: remove units
- `create-unit-occurrence-stats`: create unit occurrence statistics

#### Roadmap

- `create n-grams`
- `map units`
- `merge units right/left`
- `calculate units TF-IDF`

#### Installation

```
pip install txt-utils --user
```

#### Usage

```
txt-utils -cli
```

#### Citation

If you want to cite this repo, you can use the BibTeX-entry generated by GitHub (see About => Cite this repository).

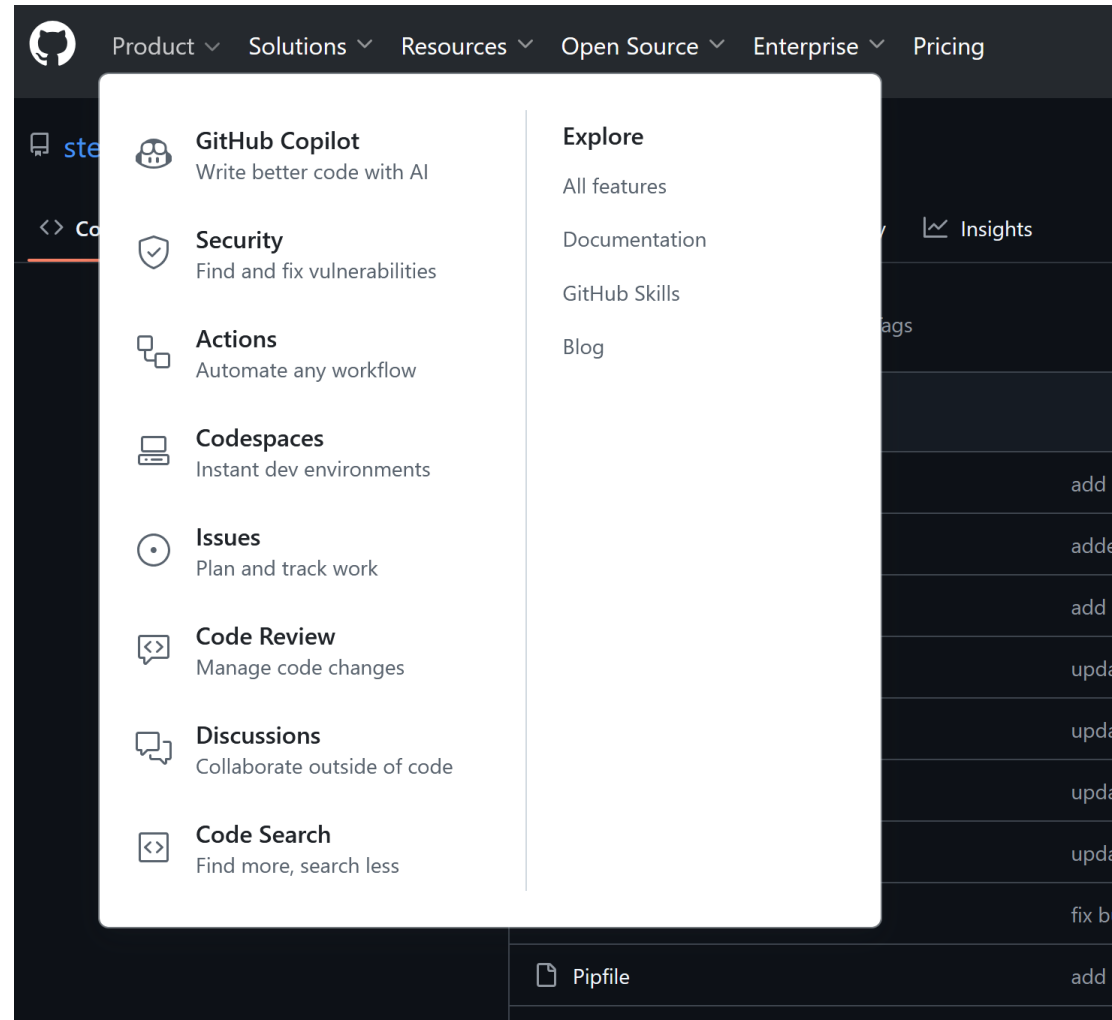
```
Taubert, S. (2024). txt-utils (Version 0.0.3) [Computer software]. https://doi.org/10.5281/zenodo.10571
```

#### Acknowledgments

Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project-ID 416228727 – CRC 1410



# Github



<https://github.com/stefantaubert/txt-utils>

# OpenAIRE - Open Access Infrastructure for Research in Europe



OpenAIRE | EXPLORE

[Search](#) [Deposit](#) [Link](#) [Data sources](#) [Funders](#)

## Discover open linked research.

A comprehensive and open dataset of research information covering 194m publications, 64m research data, 410k research software items, from 141k data sources, linked to 3m grants and 321k organizations.

All linked together through citations and semantics.

[Advanced search](#)

Search in OpenAIRE



Try browsing by:

[SUSTAINABLE DEVELOPMENT GOALS \(SDGs\)](#) →

[FIELDS OF SCIENCE \(FOS\)](#) →



<https://explore.openaire.eu/>

## Level up with the largest AI & ML community

Join over 20M+ machine learners to share, stress test, and stay up-to-date on all the latest ML techniques and technologies. Discover a huge repository of community-published models, data & code for your next project.

 Register with Google

Register with Email



### Who's on Kaggle?

#### Learners

Dive into Kaggle courses, competitions & forums.



#### Developers

Leverage Kaggle's models, notebooks & datasets.



#### Researchers

Advance ML with our pre-trained model hub & competitions.



## BirdCLEF 2024

Bird species identification from audio, focused on under-studied species in the Western Ghats, a major biodiversity hotspot in India.



[Overview](#) [Data](#) [Code](#) [Models](#) [Discussion](#) [Leaderboard](#) [Rules](#)

### Overview

#### Goal of the Competition

Birds are excellent indicators of biodiversity change since they are highly mobile and have diverse habitat requirements. Changes in species assemblage and the number of birds can thus indicate the success or failure of a restoration project. However, frequently conducting traditional observer-based bird biodiversity surveys over large areas is expensive and logistically challenging. In comparison, passive acoustic monitoring (PAM) combined with new analytical tools based on machine learning allows conservationists to sample much greater spatial scales with higher temporal resolution and explore the relationship between restoration interventions and biodiversity in depth.

For this competition, you'll use your machine-learning skills to identify under-studied Indian bird species by sound. Specifically, you'll develop computational solutions to process continuous audio data and recognize the species by their calls. The best entries will be able to train reliable classifiers with limited training data. If successful, you'll help advance ongoing efforts to protect avian biodiversity in the Western Ghats, India, including those led by V. V. Robin's Lab at IISER Tirupati.



#### Context

#### Competition Host

Cornell Lab of Ornithology



#### Prizes & Awards

\$50,000  
Awards Points & Medals

#### Participation

6,576 Entrants  
1,198 Participants  
974 Teams  
30,118 Submissions

#### Tags

Audio   Multiclass Classification  
Custom Metric



# CLEF - Conference and Labs of the Evaluation Forum



**UNLOCKING  
INFORMATION**

The CLEF Initiative promotes research, innovation, and development of *information access systems* with an emphasis on *multilingual and multimodal information with various levels of structure*. CLEF promotes research and development by providing an infrastructure for:

[Tweets by clef\\_initiative](#)

<http://www.clef-initiative.eu/>