

# Datennachnutzung in der Praxis

Angela Ariza, Esther Asef, Juliane Jacob, Andreas Mühlichen,  
Karsten Peters-von Gehlen, Hermann Schranzhofer, Ute Trautwein-Bruns

*Unter Mitwirkung weiterer Mitglieder der UAG  
Schulungen/Fortbildungen der DINI/nestor AG Forschungsdaten*



Alle Inhalte dieser Präsentation stehen unter der Lizenz [Creative Commons BY 4.0 International](https://creativecommons.org/licenses/by/4.0/), <https://doi.org/10.5281/zenodo.7568266>



# Code of Conduct



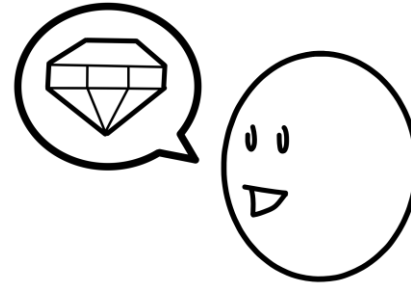
- Mikros bitte in der Regel aus
- Kameranutzung freiwillig
- Technische Probleme bitte im Chat melden
- Workshop-Du
- Beteiligt euch rege!
- Fragen bitte über die Funktion "Hand heben"



# Die UAG Schulungen/Fortbildungen

- DINI/nestor-AG Forschungsdaten
- gegründet am 5. März 2019 in Berlin
- aktuelle Mitgliederanzahl: 44
- Ziel: Weiterentwicklung und Erarbeitung von unterschiedlichen Schulungsmodellen und -materialien sowie Sammlung von Schulungsmaterialien

Orientierung				
inhaltlich	<ul style="list-style-type: none"><li>• FD-Nachnutzung<ul style="list-style-type: none"><li>• Motivation</li><li>• Such-Strategien</li><li>• Lizenzen</li><li>• Datenqualität</li><li>• Dokumentation</li></ul></li></ul>	methodisch	<ul style="list-style-type: none"><li>• Funktionen von Zoom<ul style="list-style-type: none"><li>• Breakout-Sessions, Chat</li></ul></li><li>• Einbindung externer Tools<ul style="list-style-type: none"><li>• Mentimeter</li></ul></li><li>• Unterschiedliche Sozial- und Arbeitsformen</li><li>• Fallbeispiele</li><li>• Handouts</li></ul>	konzeptionell
Fragen, Diskussion und Feedback				



1. Handzeichen: Wer hat schon FD anderer genutzt? (ja/nein)
2. Menti: Welche Datenquellen kennt ihr? (Wordcloud)
3. Menti: Welche Herausforderung seht ihr bei der Nutzung von nicht selbst erzeugten Daten? (Freitext)

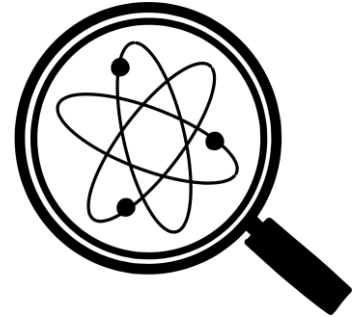


- Daten können nicht selbst generiert werden
  - Daten sind einmalig
  - Erzeugung der Daten ist sehr aufwändig/teuer
  - Langzeitstudien (z.B. sozio-ökonomisches Panel)
- Daten in neuem Kontext / mit neuer Forschungsfrage auswerten
  - Bessere Ressourcennutzung
- Daten mit eigenen kombinieren / zu etwas Größerem integrieren
  - z. B. Modelle validieren
  - Potenzial neuer Studien/Kooperationen

**Ziel: Beantworten einer Forschungsfrage**

- Berechnungen von thermischen Anlagen oder Gebäuden mit Jahressimulationen ergeben z.B.:

- Temperaturverlauf in einem Haus über das ganze Jahr
- Zeitlicher Verlauf des Energieertrages einer Solaranlage



- Wesentlich für diese Simulationen sind die Randbedingungen unter anderem werden **Wetterdaten** für ein ganzes Jahr an dem zu betrachtenden Standort benötigt.

- Es sollen Wetterdaten für den Standort Berlin gesucht werden
- Im Idealfall Stundenwerte der Temperatur [°C], Feuchte [%] und Solarstrahlung [W/m<sup>2</sup>] für ein ganzes Jahr

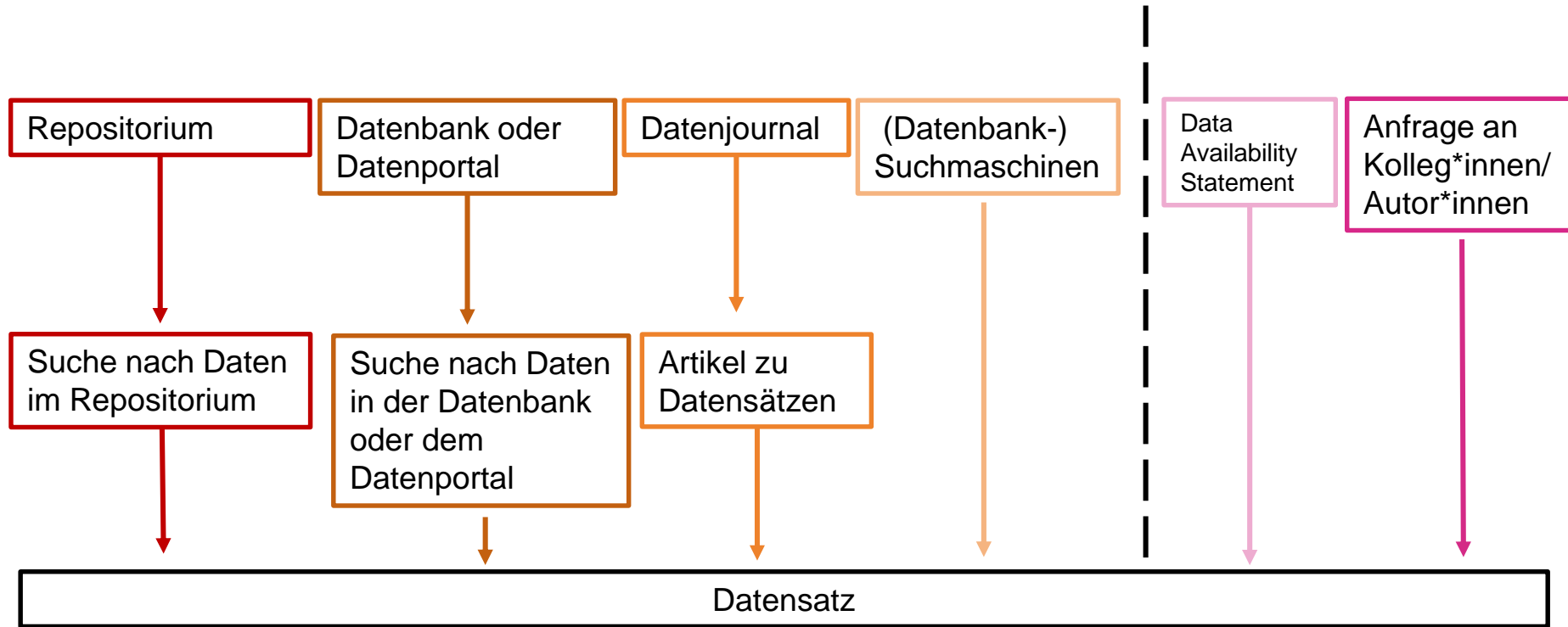
Jahresstunde [h]	Temperatur [°C]	Feuchte [%]	Solarstrahlung [W/m <sup>2</sup> ]
1	1.9	98	0
2	1.5	99	0
3	1.1	99	0
4	1	99	0
5	0.8	99	0
6	0.8	99	0
7	0.7	99	0
8	0.9	98	0
9	1.8	97	52
10	3.2	94	142
11	4.7	87	245
12	6.3	82	328
13	5.8	84	303
14	5.8	83	185
15	5.8	83	108



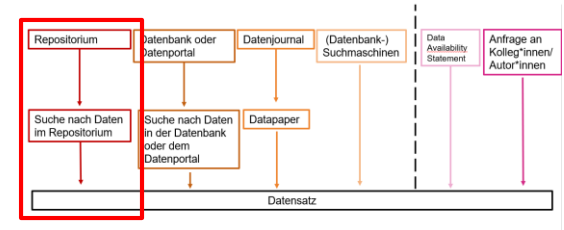


# Übung Wetterdatensuche - Ablauf

- 10 min Zeit
- Fragestellungen:
  - Findet man Daten und wo?
  - Kann man die Daten auch herunterladen?
  - Sind die Daten verständlich und brauchbar?
  - Kann/darf man die Daten wirklich nutzen?
- Ergebnisse und ggf. Links in den Chat schreiben
- Fazit

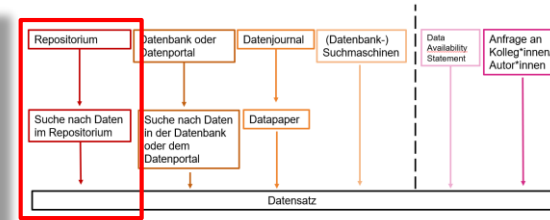


Repositorien finden: [re3data.org](https://re3data.org)



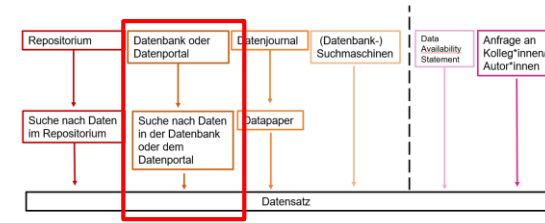
The screenshot shows the re3data.org search interface. The search term 'education sciences' is entered in the search bar. The results list includes:

- National Center for Education Statistics (NCES)**: A result for 'Education Sciences' with a description: "The National Center for Education Statistics (NCES) is responsible for collecting and analyzing data related to education, including assessing the performance of students from early childhood through secondary education as well as the literacy level of adults and post-secondary education surveys. Users can access data on public and private schools as well as public libraries and a college navigator tool containing information on over 7,000 post-secondary institutions." This description is highlighted with a red box.
- Unified-District Information System for Education (U-DISE)**: A result for 'Education Sciences' with a description: "It is a statistical system developed for collection, computerization, analysis and use of educational and allied data for planning."



## Fachübergreifendes Verzeichnis

- DBIS [Fachübersicht](#)



DATENBANK-INFOSYSTEM (DBIS)  
Gesamtbestand in DBIS

FACHÜBERSICHT

FACHGEBIETE	ANZAHL
Allgemein / Fachübergreifend	2785
Allgemeine und vergleichende Sprach- und Literaturwissenschaft	774
Anglistik, Amerikanistik	726
Archäologie	298
Architektur, Bauingenieur- und Vermessungswesen	506
<b>Biologie</b>	<b>757</b>
Chemie	536
Elektrotechnik, Mess- und Regelungstechnik	740

DATENBANK-INFOSYSTEM (DBIS)  
Gesamtbestand in DBIS

FACHGEBIET: BIOLOGIE

AUSWAHL

Sortierung der Ergebnisse

alphabetisch  
alphabetisch  
nach Datenbanktyp

GESAMTANGEBOT (780 TREFFER)

10 Jahre Digitale Bibliothek: Jubiläumsausgabe 2 / DVD 3	CD-ROM/DVD
2000 Pilze	CD-ROM/DVD
AAS Open Research	frei im Web Start
Academic Search Complete (via EBSCO Host)	WWW

DATENBANK-INFOSYSTEM (DBIS)  
Gesamtbestand in DBIS

FACHGEBIET: BIOLOGIE

AUSWAHL

Sortierung der Ergebnisse

Rubrik

Bestandsverzeichnis (10)

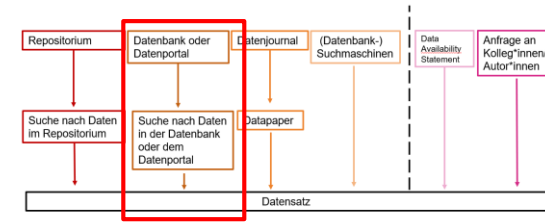
- Bilddatenbank (41)
- Biographische Datenbank (7)
- Buchhandelsverzeichnis (2)
- Dissertationsverzeichnis (3)
- Fachbibliographie (154)
- Faklendenbank (170)
- National-, Regionalbibliographie (2)
- Portal (75)
- Volltextdatenbank (258)
- Wörterbuch, Enzyklopädie, Nachschlagewerk (192)
- Adress- und Firmenverzeichnis (7)

ADRESS- UND FIRMENVERZEICHNIS (7 TREFFER)

International Directory for Bio-based Businesses	frei im Web Start
Kompetenzatlas Nanotechnologie in Deutschland	frei im Web Start

## Fachübergreifendes Verzeichnis

- DBIS [Fachübersicht](#)



FAKTENDATENBANK (170 TREFFER) <span style="float: right;">ZUGANG</span>
<ul style="list-style-type: none"> <li>• FAKTENSAMMLUNGEN (FORMAL UND INHALTLICH ERSCHLOSSENE PRIMÄRINFORMATIONEN), Z.B. STATISTISCHE DATEN, CHEMISCHE FORMELN, FIRMEN- UND PRODUKTDATEN</li> </ul>
Plantwise Knowledge Bank <span style="float: right;">frei im Web</span>
Primal Pictures Interactive Anatomy <span style="float: right;">WWW</span>
PRINTS <span style="float: right;">frei im Web</span>
Projektdatenbank PLANT 2030 <span style="float: right;">frei im Web</span>
<b>Protein Data Bank</b> <span style="float: right;">frei im Web</span>
PubChem <span style="float: right;">frei im Web</span>
QueRBeet <span style="float: right;">frei im Web</span>



**DATENBANK-INFOSYSTEM (DBIS)**  
Gesamtbestand in DBIS

DETAILANSICHT ◀ ▶

**PROTEIN DATA BANK**

Bibliothek(en) mit Bestandsnachweis:  
4 Fh-Standorte in Kärnten: Klagenfurt, Villach, Spittal An Der D

Go!

Weitere Titel: PDB, Protein Data Bank, RCSB Protein Data Bank

Recherche starten: <https://www.rcsb.org/pdb/home/home.do>

Verfügbar: Freier Zugang

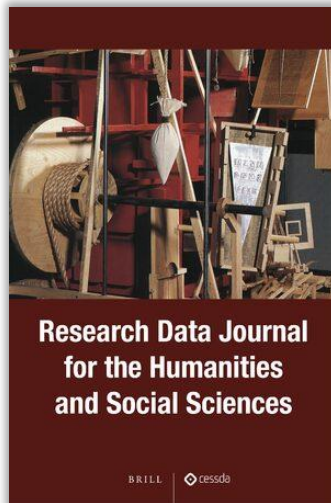
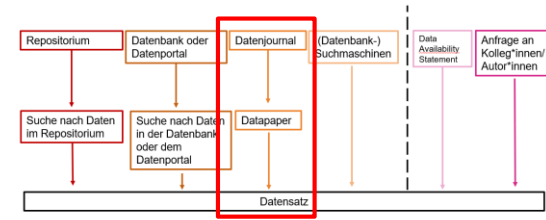
Inhalt: Protein-Datenbank, die von mehreren US-Forschungsinstitutionen erstellt wird. Inhalt ist die chemische Struktur von Proteinen.

Fachgebiete: Biologie, Chemie, Verfahrenstechnik, Biotechnologie, Lebensmitteltechnologie

Schlagerwörter: Proteine

## Liste mit [Datenjournalen](#)

- publizieren ausführliche Beschreibung der Daten
- meist peer-reviewed
- Verlinkung zum Datensatz



 [A Data Set for US Horror Film Trailers](#)

Arts and Media

Autor:in: [Nick Redfern](#)

Seiten: 1–12

Online-Publikationsdatum: 09 Nov 2021

 [PDF herunterladen](#)





**A Data Set for US Horror Film Trailers**

Arts and Media  
in *Research Data Journal for the Humanities and Social Sciences*  
Autorin: [Nick Redfern](#)  
Online-Publikationsdatum: 09. Nov. 2021

Abstract Volltext PDF Metadaten Quellen Abbildungen Kennzahlen

**Abstract**

This article presents a new data set comprising audio, colour, motion, and shot length data of trailers for the fifty highest grossing horror films at the US box office from 2011 to 2015. This data set is one of the few available for computational film analysis that includes data on multiple elements of film style and is the only existing data set for motion picture trailers suitable for formal analyses. **Data is stored in .osx files available under a Creative Commons Attribution 4.0 International license on Zenodo: [www.doi.org/10.5281/zenodo.4479068](https://www.doi.org/10.5281/zenodo.4479068)**

Schlüsselwörter: **computational film analysis; horror cinema; film trailers; sound; colour; editing; motion**

Online publication date: 09-11-2021

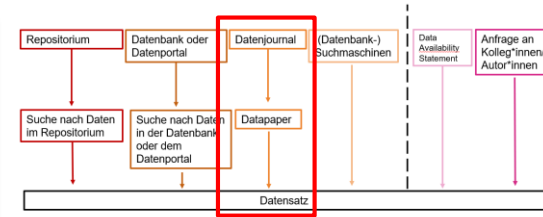
- Related data set "US Horror Trailers Data Set" with DOI [www.doi.org/10.5281/zenodo.4479068](https://www.doi.org/10.5281/zenodo.4479068) in repository "Zenodo"

**1. Introduction**

David Bordwell describes film style as "a film's systematic and significant use of the techniques of the medium ... Style is, minimally, the texture of the film's images and sounds, the result of choices made by filmmaker(s) in particular historical circumstances" (1997, p. 4). Film style can be divided into four broad categories: mise-en-scène, cinematography, editing, and audio (see Figure 1). Our purpose in analysing the style of a film is "to explain why an individual motion picture is the way it is: why it has the elements of style it does and why they stand in the relations that they do" (Carroll, 2009, p. 268). Computational film analysis lies squarely within the digital humanities, and aims to

Open Access  
PDF herunterladen  
Zitierung herunterladen  
Berechtigungen erhalten

Research Data Journal for the Humanities and Social Sciences







# Datenjournale



zenodo Search Upload Communities Log in Sign up

January 29, 2021 Dataset Open Access

## US\_Horror\_Trailers\_Data\_Set

Redfern, Nick

This data set contains audio, colour, editing, and motion data for trailers for the fifty highest grossing horror films at the US box office from 2011 to 2015.

All trailers were pre-processed to trim MPAA tag screens and YouTube channel promotional materials, and cropped to remove letterbox blanking.

The sample is described in the file `US_Horror_Trailers_Sample_Summary.csv`, which contains the title of the film promoted, the URL for the trailer on YouTube (all URLs were correct as of 27 January 2021), the dimensions of the video file after pre-processing, the number of frames after pre-processing, the native frame rate, and the running time of a trailer after pre-processing in seconds.

For each trailer the following information is available:

- Audio: the time contour of the normalized aggregated power envelope. One csv file per trailer with the naming convention: `*_audio.csv`
- Colour: three unnamed columns containing the average colour of each frame as an RGB triplet. One csv file per trailer with the naming convention: `*_rgb.csv`
- Motion: total magnitude and horizontal and vertical pixel displacement between consecutive frames. One csv file per trailer with the naming convention: `*_motion.csv`
- Editing: Shot length data in seconds. Shot length data is contained in the csv file `US_Horror_Trailers_SL_Data.csv`, with one column per data.

All variables are time ordered from frame/shot/window 1 to  $n$  even when no explicit time variable is provided.

A guide to visualising the data in this data set is available online here: [https://rpubs.com/nr62\\_rp33/visualizing\\_trailers](https://rpubs.com/nr62_rp33/visualizing_trailers).

A data paper describing this data set can be accessed at: Redfern, N. (2021) A Data Set for US Horror Film Trailers, *Research Data Journal for the Humanities, and the Social Sciences* 6 (1): 1-12. <https://doi.org/10.1163/24523666-bja10017>

164 views 40 downloads

Indexed in OpenAIRE

Publication date: January 29, 2021

DOI: [10.5281/zenodo.4479067](https://doi.org/10.5281/zenodo.4479067)

Keyword(s): Film, Cinema, Computational Film Analysis, Film sound, Film editing, Film colour, Motion, Horror cinema, Film trailers

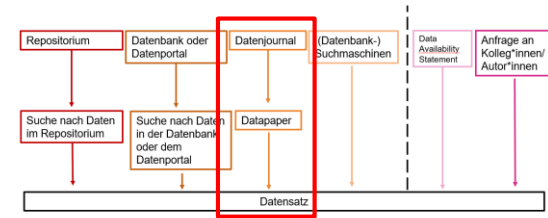
License (for files): Creative Commons Attribution 4.0 International

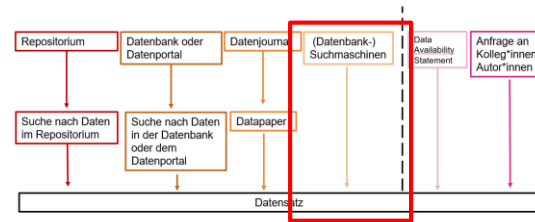
Preview US\_Horror\_Trailers\_Audio\_Data.zip

- US\_Horror\_Trailers\_Audio\_Data
  - A\_Haunted\_House\_2\_audio.csv 118.5 kB
  - A\_Haunted\_House\_audio.csv 230.7 kB

Versions: Version 1.0 (Jan 29, 2021) 10.5281/zenodo.4479068

Cite all versions? You can cite all versions by using the DOI 10.5281/zenodo.4479067. This DOI represents all





Service	Link	Betreiber	Indices
BASE	<a href="http://www.base-search.net">www.base-search.net</a>	Universitätsbibliothek Bielefeld	Wissenschaftliche Web-Dokumente, darunter auch Forschungsdaten
B2FIND	<a href="http://b2find.eudat.eu/">http://b2find.eudat.eu/</a>	EUDAT CDI	Sammlungen der Datenzentren von EUDAT und weiterer europäischer Repositorien
DataCite Metadata Search	<a href="https://search.datacite.org/">https://search.datacite.org/</a>	DataCite	Mit DOI erschlossene Daten bzw. digitale Objekte (auch Textpublikationen)
Dataset Search	<a href="https://datasetsearch.research.google.com/">https://datasetsearch.research.google.com/</a>	Google	Datensets

**DataCite Search**
Works People Repositories Members Support ➔ Sign in

Search

**10,682 Works**

**Berlin Climate Record - Inner City - 1766-1934**  
 Christopher Kadow, Bianca Wentzel, Ilona Jaekel, Marcus Bahlo & Ulrich Cubasch  
 Collection Of Datasets published via PANGAEA - Data Publisher for Earth & Environmental Science  
 Berlin (Germany) has one of the longest climate records in the world (Cubasch and Kadow, 2011). In the 17th century mainly the family of the astronomer Kirch started measuring the temperature and wrote down general weather patterns. In the beginning of the 18th century the measurements became more regular including multiple measures per day. Over time, more inner city stations appeared but also disappeared. Therefore, this Berlin Climate Record is a moving station, but representable...

i No citations were reported. No usage information was reported.

<https://doi.org/10.1594/pangaea.870862> Cite

**Berlin Climate Record - Inner City - 1766-1934**  
 Christopher Kadow, Bianca Wentzel, Ilona Jaekel, Marcus Barlow & Ulrich Cubasch  
 Dataset published via Freie Universität Berlin  
 Berlin (Germany) has one of the longest climate records in the world (Cubasch and Kadow, 2011). In the 17th century mainly the family of the astronomer Kirch started measuring the temperature and wrote down general weather patterns. In the beginning of the 18th century the measurements became more regular including multiple measures per day. Over time, more inner city stations appeared but also disappeared. Therefore, this Berlin Climate Record is a moving station, but representable...

**Registration Year**

<input type="checkbox"/> 2022	2,685
<input type="checkbox"/> 2021	2,877
<input type="checkbox"/> 2020	2,098
<input type="checkbox"/> 2019	1,595
<input type="checkbox"/> 2018	839
<input type="checkbox"/> 2017	168
<input type="checkbox"/> 2016	274
<input type="checkbox"/> 2015	67
<input type="checkbox"/> 2014	6
<input type="checkbox"/> 2013	2

**Resource Types**

<input type="checkbox"/> Dataset	9,832
<input type="checkbox"/> Text	598
<input type="checkbox"/> Journal Article	69
<input type="checkbox"/> Audiovisual	21
<input type="checkbox"/> Other	14
<input type="checkbox"/> Software	12
<input type="checkbox"/> Collection	11
<input type="checkbox"/> Event	11
<input type="checkbox"/> Conference Paper	9
<input type="checkbox"/> Report	8

The screenshot displays the B2FIND data catalogue interface. At the top left, there are logos for B2FIND and EUDAT. The navigation menu includes: DATA CATALOGUE, REPOSITORIES, COMMUNITIES, FOR PROVIDERS, FOR USERS, and ABOUT. The main heading is "Datasets".

A search bar contains the text "Climate Berlin" and a search icon. Below the search bar, it states "377 datasets found for 'Climate Berlin'" and "Order by: Relevance".

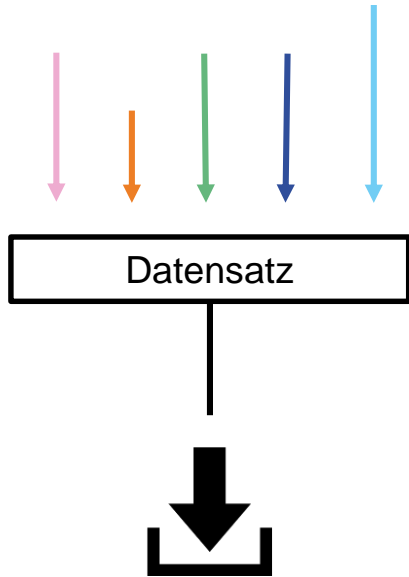
On the left side, there is a "Spatial Coverage" section with a map showing a heatmap of Europe. Below the map is a list of filter categories, each with a dropdown arrow. A red box highlights this list:

- Temporal Coverage
- Publication Year
- Repositories
- Communities
- Keywords
- Creator
- Instrument
- Discipline
- Language

The main content area shows the following dataset entries:

- Berlin Climate Record - Inner City - 1766-1934**  
Berlin (Germany) has one of the longest climate records in the world (Cubasch and Kadow, 2011). In the 17th century mainly the family of the astronomer Kirch started measuring...
- Atmosphere output of the Bergen Climate Model experiment E76**  
80 years CMIP2 1% per year compound CO2 increase experiment starting from present-day climate state. BCM1.0 (2002): atmosphere: ARPEGE V3 T63L31 (cycle 22); ocean: NERSC-MICOM...
- Atmosphere output of the Bergen Climate Model experiment E99**  
Freshwater sensitivity experiment based on pre-industrial control simulation BCM1-E75 (Furevik et al. 2003, doi:10.1007/s00382-003-0317-5). In E99 the Arctic river runoff is...
- Ocean output of the Bergen Climate Model experiment L12**  
150 years historical run from 1850-1999, including transient external and anthropogenic forcings (e.g. total solar irradiance, volcanic eruptions, tropospheric aerosols and...

# Zugang zum Datensatz

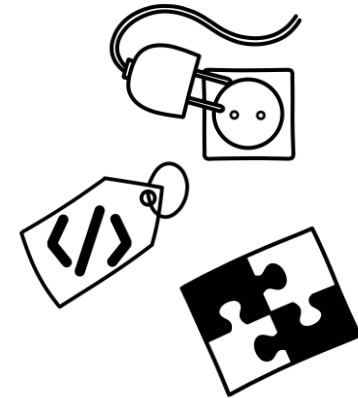


Wenn Daten gefunden wurden, sind diese auch zugänglich?

Verschiedene Zugangskategorien am Beispiel [SowiDataNet|datorium](#)

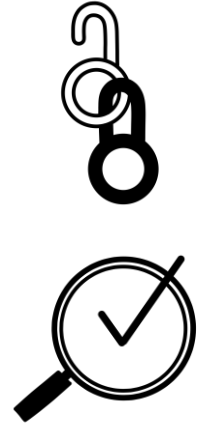
- **Freier Zugang (ohne Registrierung):** Die Forschungsdaten sind unter Beachtung der Nutzungsbedingungen ohne weitere Einschränkung für alle Nutzenden zugänglich.
- **Freier Zugang (mit Registrierung):** Die Forschungsdaten sind unter Beachtung der Nutzungsbedingungen ohne weitere Einschränkung für alle registrierten Nutzenden zugänglich.
- **Eingeschränkter Zugang:** Die Forschungsdaten sind nur eingeschränkt verfügbar. Wünschen Nutzende einen Zugang zu Forschungsdaten mit dieser Zugangsklasse, müssen sie die Zustimmung der Datengebenden einholen. SowiDataNet|datorium erleichtert diesen Prozess: Beim Versuch des Downloads wird ein Mailformular generiert, mit dem die Datennutzung bei einem Kurator/einer Kuratorin beantragt werden kann. Nach Rücksprache mit dem Datengeber/der Datengeberin veranlasst diese/r die Freischaltung der Forschungsdaten für die beantragende Person.

- Daten wurden gefunden und sind zugänglich
- Sind folgende Punkte erfüllt?
  - **Offenes, nicht proprietäres Format**
  - **Formatumwandlung verlustfrei möglich**
  - **Standardisierte Vokabulare, Wertebereiche...**
  - **Ggf. offener Zugang zu Software**
- Dann sind die Daten wahrscheinlich **interoperabel**

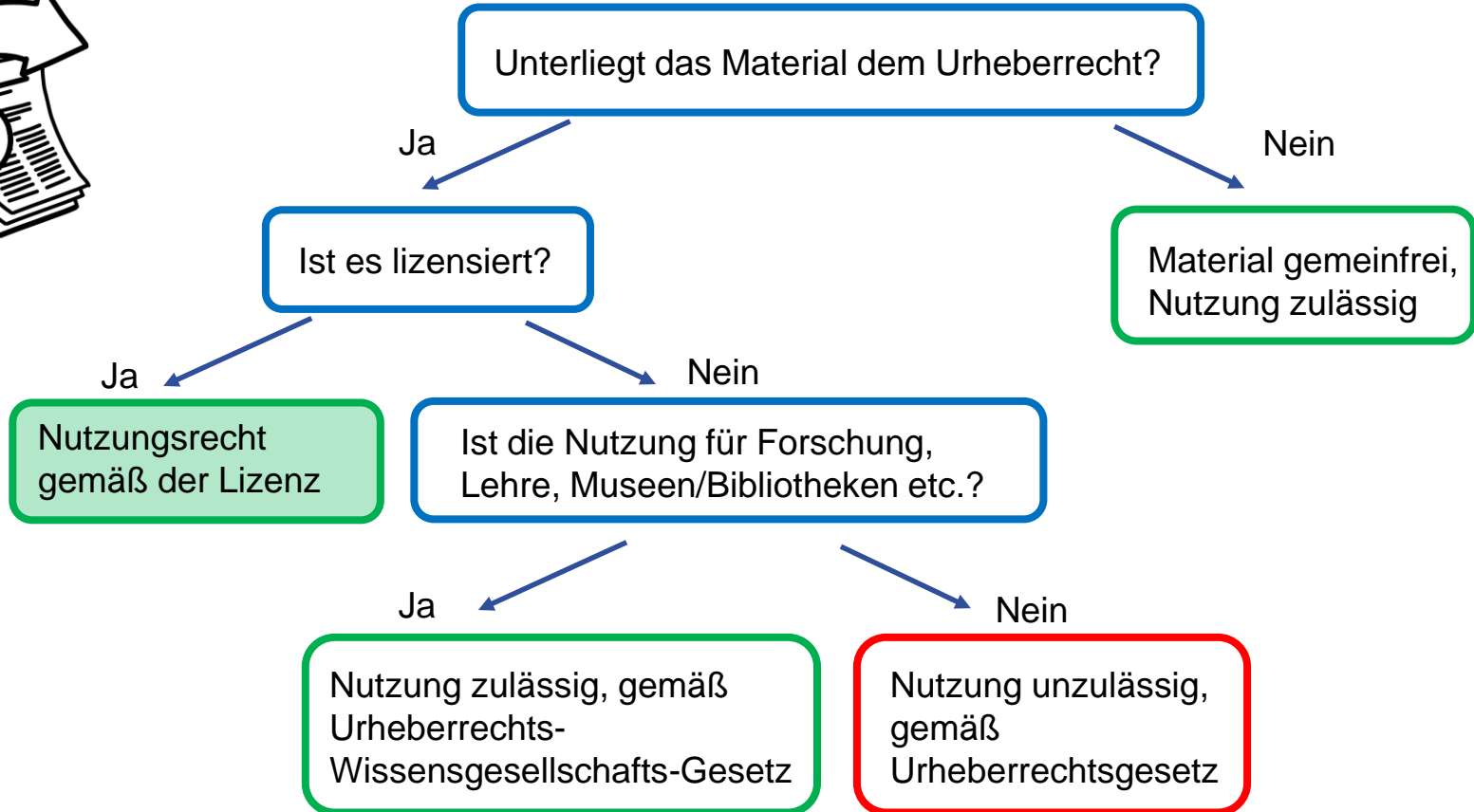


# Wiederverwendbarkeit von Daten

- Daten wurde gefunden, sind zugänglich und interoperabel
- Sind folgende Punkte beantwortet?
  - **Nutzungsrechte – Stichwort „Lizenzen“**
  - **Dokumentation**
  - **Datenqualität**
- Dann sind die Daten wahrscheinlich **wiederverwendbar**



# Rechtliche Aspekte der Nachnutzung





- Lizenzen regeln die Bedingungen der Nutzung von veröffentlichten Daten
- Creative Commons-Lizenzen: Standardnutzungsverträge zwischen Nutzenden und Rechteinhabern
- <https://creativecommons.org>



Namensnennung (BY)



nicht-kommerziell (NC)



keine Derivate (ND)



Weitergabe unter gleichen Lizenzbedingungen (SA)

# Nutzungsrechte Lizenzangaben in den Metadaten

**Citation:** **Kadow, Christopher; Wentzel, Bianca; Jaekel, Ilona; Bahlo, Marcus; Cubasch, Ulrich (2017):** Berlin Climate Record - Inner City - 1766-1934. *Institut für Meteorologie, Freie Universität Berlin, PANGAEA*, <https://doi.org/10.25418/PANGAEA.870862>


Always quote citation above when using data: You can download the citation in several formats below.

[RTS Citation](#)
[BisToX Citation](#)
[S Copy Citation](#)
[Facebook](#)
[Twitter](#)
[Show Map](#)
[Google Earth](#)

**Abstract:** Berlin (Germany) has one of the longest climate records in the world (Cubasch and Kadow, 2011). In the 17th century mainly the family of the astro patterns. In the beginning of the 18th century the measurements became more regular including multiple measures per day. Over time, more inner is a moving station, but representable for the inner city of Berlin. With these datasets we reactivate the long inner city climate record in several var datasets on the very rare daily time frequency.

**Related to:** **Cubasch, Ulrich; Kadow, Christopher (2011):** Global climate change and aspects of regional climate change in the Berlin-Brandenburg region. [erde.org/index.php/die-erde/article/download/40/35](https://www.erde.org/index.php/die-erde/article/download/40/35)

**Coverage:** *Latitude:* 52.506000 \* *Longitude:* 13.397440  
*Date/Time Start:* 1766-01-01T12:00:00 \* *Date/Time End:* 1934-12-31T12:00:00

**License:**  Creative Commons Attribution-ShareAlike 3.0 Unported (CC-BY-SA-3.0)

**Size:** 2 datasets

January 29, 2021 Dataset Open Access

## US\_Horror\_Trailers\_Data\_Set

Redfern, Nick

This data set contains audio, colour, editing, and motion data for trailers for the fifty highest grossing horror films at the US box office from 2011 to 2015.

All trailers were pre-processed to trim MPAA tag screens and YouTube channel promotional materials, and cropped to remove letterbox blanking.

The sample is described in the file US\_Horror\_Trailers\_Sample\_Summary.csv, which contains the title of the film promoted, the URL for the trailer on YouTube (all URLs were correct as of 27 January 2021), the dimensions of the video file after pre-processing, the number of frames after pre-processing, the native frame rate, and the running time of a trailer after pre-processing in seconds.

For each trailer the following information is available:

- Audio: the time contour of the normalized aggregated power envelope. One csv file per trailer with the naming convention: \*\_audio.csv
- Colour: three unnamed columns containing the average colour of each frame as an RGB triplet. One csv file per trailer with the naming convention: \*\_rgb.csv
- Motion: total magnitude and horizontal and vertical pixel displacement between consecutive frames. One csv file per trailer with the naming convention: \*\_motion.csv
- Editing: Shot length data in seconds. Shot length data is contained in the csv file US\_Horror\_Trailers\_SL\_Data.csv, with one column per data.

All variables are time ordered from frame/shot/window 1 to  $n$  even when no explicit time variable is provided.

A guide to visualising the data in this data set is available online here: [https://rpubs.com/hr62\\_rp33/visualizingtrailers](https://rpubs.com/hr62_rp33/visualizingtrailers).

A data paper describing this data set can be accessed at: Redfern, N. (2021) A Data Set for US Horror Film Trailers, *Research Data Journal for the Humanities, and the Social Sciences* 6 (1): 1-12. <https://doi.org/10.1163/24523666-bja10017>

165 views 40 downloads [See more details...](#)

Indexed in **OpenAIRE**

**Publication date:** January 29, 2021

**DOI:** [10.5281/zenodo.4479063](https://doi.org/10.5281/zenodo.4479063)

**Keyword(s):** [Film editing](#) [Film colour](#) [Motion](#) [Horror cinema](#)

**License (for files):** [Creative Commons Attribution 4.0 International](#)

**Versions**  
Version 1.0 Jan 29, 2021

# Nutzungsrechte Policy der Infrastruktur

## Beispiel [RCSB Protein Data Bank](#)

### Usage Policies

The [wwPDB policy](#) states that data files contained in the PDB archive are available under [the CC0 1.0 Universal \(CC0 1.0\) Public Domain Dedication](#).

All data provided by RCSB PDB programmatic APIs are available under the same license.

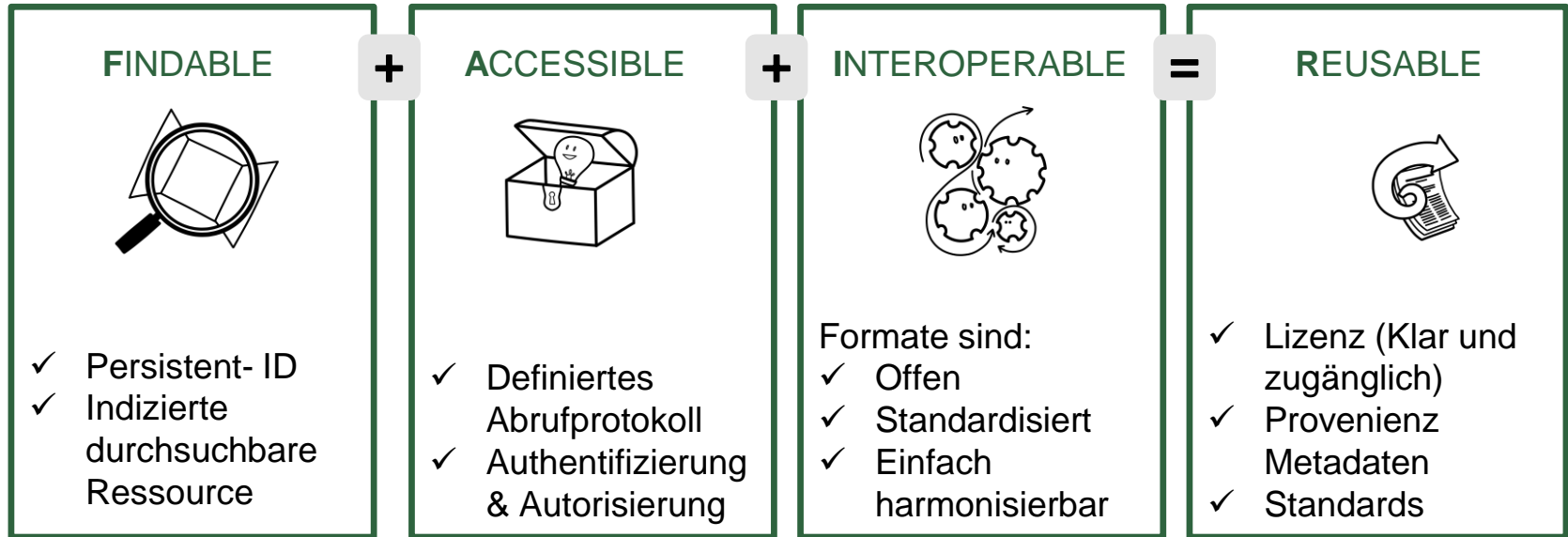
All data are presented without any warranties about the work; RCSB PDB disclaims liability for all uses of the work, to the fullest extent permitted by applicable law.

Users of PDB data are encouraged to [attribute the original authors of the PDB structure data](#) where possible.

[Full RCSB PDB usage policies](#) are available.

# Datenqualität: FAIRe Daten

## FAIR (META)DATEN Prinzipien



... für Menschen und Maschinen

Übersetzt und angepasst von: [Wilbrandt \(2021\)](#)

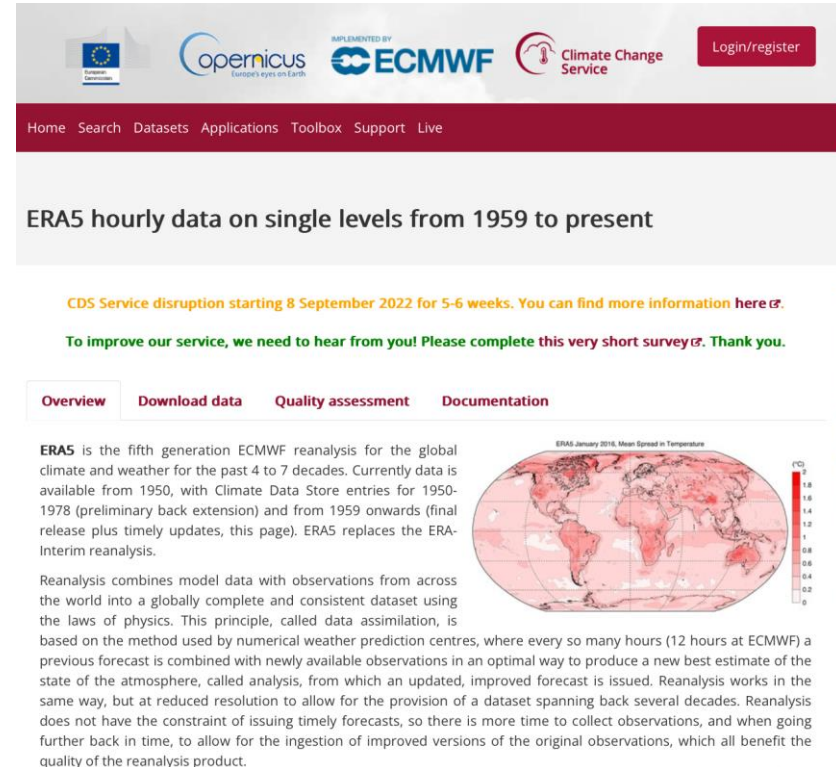
## Übung:

Welche Aspekte der FAIR-Prinzipien sind auf dieser Landingpage zu finden?

- [Landingpage Datensatz](#)

Antworten bitte in den **Chat** posten:

- F: „das Merkmal“
- A: „noch ein Merkmal“
- usw.



The screenshot shows the ERA5 landing page with the following elements:

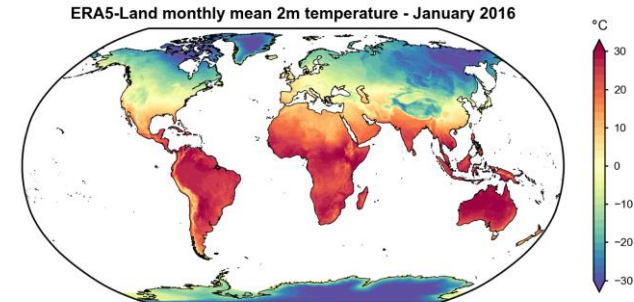
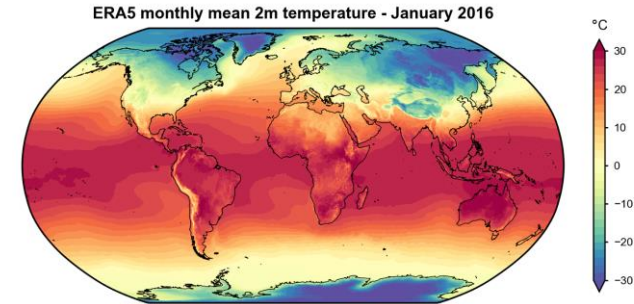
- Header:** Logos for the European Union, Copernicus (Europe's eyes on Earth), ECMWF (implemented by), and Climate Change Service. A "Login/register" button is on the right.
- Navigation:** A dark red bar with links for Home, Search, Datasets, Applications, Toolbox, Support, and Live.
- Main Title:** "ERA5 hourly data on single levels from 1959 to present".
- Alerts:**
  - Yellow text: "CDS Service disruption starting 8 September 2022 for 5-6 weeks. You can find more information here ↗".
  - Green text: "To improve our service, we need to hear from you! Please complete this very short survey ↗. Thank you."
- Navigation Tabs:** Overview (selected), Download data, Quality assessment, Documentation.
- Text Content:**

**ERA5** is the fifth generation ECMWF reanalysis for the global climate and weather for the past 4 to 7 decades. Currently data is available from 1950, with Climate Data Store entries for 1950-1978 (preliminary back extension) and from 1959 onwards (final release plus timely updates, this page). ERA5 replaces the ERA-Interim reanalysis.

Reanalysis combines model data with observations from across the world into a globally complete and consistent dataset using the laws of physics. This principle, called data assimilation, is based on the method used by numerical weather prediction centres, where every so many hours (12 hours at ECMWF) a previous forecast is combined with newly available observations in an optimal way to produce a new best estimate of the state of the atmosphere, called analysis, from which an updated, improved forecast is issued. Reanalysis works in the same way, but at reduced resolution to allow for the provision of a dataset spanning back several decades. Reanalysis does not have the constraint of issuing timely forecasts, so there is more time to collect observations, and when going further back in time, to allow for the ingestion of improved versions of the original observations, which all benefit the quality of the reanalysis product.
- Figure:** A world map titled "ERA5 January 2018, Mean Spread in Temperature". The map shows a color scale from 0 to 1.8, with red indicating higher values. The map shows higher spread values (red) in the mid-latitude regions and lower values (white/yellow) in the tropics.

## Daten: ERA5 Reanalysen für den Zeitraum von 1959 bis heute

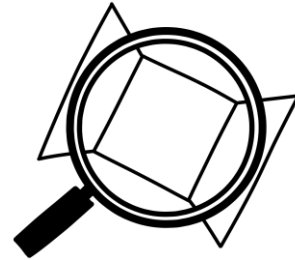
- Reanalysen stellen die genaueste Abschätzung des Zustandes der Atmosphäre in der Vergangenheit dar
- ERA5 ist der aktuellste, vom ECMWF (*European Center for Medium-Range Weather Forecasts*) verfügbare globale Reanalysedatensatz
- ERA5 Daten gibt es für den Zeitraum 1959 bis “heute minus 3 Monate”
- ERA5 hat eine räumliche und zeitliche Auflösung von  $1/4^\circ \times 1/4^\circ$  (ca.  $25 \times 25 \text{ km}^2$ ) und 1h



Datensatz : (“ERA5 hourly data on single levels from 1959 to present”)

## Persistent Identifier (PID)?

- DOI: [10.24381/cds.adbb2d47](https://doi.org/10.24381/cds.adbb2d47)
- Ausgiebige Metadaten auf der Landingpage
- Findbar in zB Google Dataset Search



Datensatz : (“ERA5 hourly data on single levels from 1959 to present”)

▼ Zuletzt aktualisiert    ▼ Download-Format    ▼ Nutzungsrechte    ▼ Thema    Kostenlos    Gespeicherte Datensätze

Mehr als 100 Datensätze gefunden

ECMWF Reanalysis v5  
ecmwf.int  
grib  
Aktualisiert: Feb 28, 2020

ERA5 Reanalysis  
rda.ucar.edu  
search.datacite.org  
Aktualisiert: Sep 5, 2017

ECMWF ERA5 Reanalysis  
registry.opendata.aws  
Aktualisiert: Jul 12, 2018

ERA5 Reanalysis (0.25 Degree Latitude-Longitude Grid)  
rda.ucar.edu  
Aktualisiert: May 13, 2019

ECMWF Reanalysis v5  
ERA5

Explore at: [Copernicus Climate Data Store](#)

338 wissenschaftliche Artikel zitieren diesen Datenpool ([In Google Scholar ansehen](#))

grib

**Datensatz aktualisiert**  
Feb 28, 2020

**Datenpool erstellt und bereitgestellt von**  
[European Centre for Medium-Range Weather Forecasts](#)

**Lizenz**  
<http://apps.ecmwf.int/datasets/licences/copernicus>

**Abgedeckter Zeitraum**  
Jan 1950 - Jetzt

**Beschreibung**  
ERA5 is the fifth generation ECMWF atmospheric reanalysis of the global climate covering the period from January 1950 to present. ERA5 is produced by the Copernicus Climate Change Service (C3S) at ECMWF. ERA5 provides hourly estimates of a large

DOI auf der Landingpage

## References

Citation

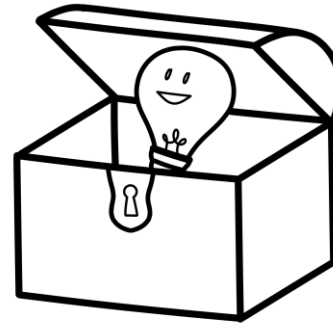
Acknowledgement

DOI: [10.24381/cds.adbb2d47](https://doi.org/10.24381/cds.adbb2d47)



(Meta)Daten zugreifbar?

- Metadaten ohne Login zugreifbar
- Datensubset spezifisch im Webinterface auswählbar
- Download mit Login
- Download via HTTP oder Python API (sinnvoll für große Datenmengen)



## Datenauswahl

Overview **Download data** Quality assessment Documentation

Clear all

Product type

Reanalysis  Ensemble mean  Ensemble members  Ensemble spread

Select all Clear all

Variable ?

At least one selection must be made

▼ Popular

<input type="checkbox"/> 10m u-component of wind	<input type="checkbox"/> 10m v-component of wind
<input type="checkbox"/> 2m dewpoint temperature	<input type="checkbox"/> 2m temperature
<input type="checkbox"/> Mean sea level pressure	<input type="checkbox"/> Mean wave direction
<input type="checkbox"/> Mean wave period	<input type="checkbox"/> Sea surface temperature
<input type="checkbox"/> Significant height of combined wind waves and swell	<input type="checkbox"/> Surface pressure
	<input type="checkbox"/> Total precipitation

Select all

▶ Temperature and pressure

▶ Wind

▶ Mean rates



## Login für Download

Format

GRIB  NetCDF (experimental)

Clear all

Show API request Show Toolbox request Login/register to submit request

## Download via HTTP

All Queued In progress Failed Unavailable Complete

Auto refreshed : 10:45:06 Delete selected

Product	Submission date	End date	Duration	Size	Status	
▶ ERA5 hourly data on single levels from 1959 to present	2022-11-02 10:44:55		0:00:11		In progress	<input type="checkbox"/>



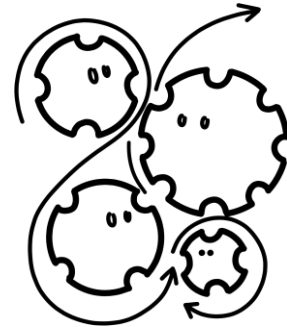
All Queued In progress Failed Unavailable Complete

Auto refreshed : 10:45:58 Delete selected

Product	Submission date	End date	Duration	Size	Status	
▶ ERA5 hourly data on single levels from 1959 to present	2022-11-02 10:44:55	2022-11-02 10:45:51	0:00:55	1.5 GB	<span>Download</span>	<input type="checkbox"/>

## Standards

- disziplinspezifisches Datenformat (netCDF, GRIB)
- Georeferenzierte Daten
- Physikalische Einheiten nach SI
- Etablierter (Meta-) Datenstandard: [CF-1.6](#), (kontrolliertes Vokabular)



Nutzungsbedingungen?  
Kontextinformationen  
/Dokumentation?

- offene Lizenz
- Umfangreiche Kontextinformationen im Reiter “Dokumentation”
- Dokumentation von Qualitätsparametern



## Zitationsvorgabe

Citation ×

Hersbach, H., Bell, B., Berrisford, P., Biavati, G., Horányi, A., Muñoz Sabater, J., Nicolas, J., Peubey, C., Radu, R., Rozum, I., Schepers, D., Simmons, A., Soci, C., Dee, D., Thépaut, J.-N. (2018): ERA5 hourly data on single levels from 1959 to present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS). (Accessed on < DD-MMM-YYYY >), 10.24381/cds.adbb2d47

[Copy to clipboard](#) [Close](#)

## Dokumentation

[Overview](#) [Download data](#) [Quality assessment](#) [Documentation](#)

- [ERA5 data documentation](#) [↗](#)  
Detailed information relating to the ERA5 data archive can be found in the web link above.
- [The ERA5 global reanalysis: Preliminary extension to 1950](#) [↗](#)  
Journal article describing the ERA5 preliminary extension.
- [The ERA5 global reanalysis](#) [↗](#)  
Journal article describing ERA5.
- [Renamed variable: form ocean waves 10m wind to ocean surface stress equivalent 10m neutral wind](#) [↗](#)  
The reason for the change was a parameter name clash between variables in ERA5 wind and ERA5 ocean waves.

## Qualitätsbewertung

[Lacagnina et al.\(2022\)](#)

Overview

Download data

Quality assessment

Documentation

This is a new feature, work in progress. Should any inconsistency be found, please report to <https://support.ecmwf.int>



The CDS datasets are assessed by the Evaluation and Quality Control (EQC) function of C3S independently of the data supplier. EQC encompasses a framework of processes aimed to assure technical and scientific quality harmonized across all dataset types available through the CDS. During the EQC process, the documentation provided with the dataset is scrutinized and data are checked for usability and reliability.

Variable:

2m temperature ✕

▼ Variable: 2m temperature

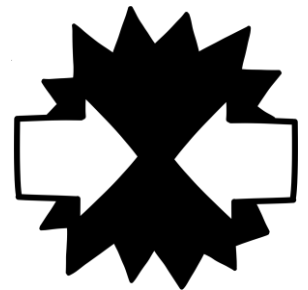
[external link icon](#) Last updated on 21/09/2021

INTRODUCTION	USER DOCUMENTATION	ACCESS	INDEPENDENT ASSESSMENT
Dataset overview	User guide	Toolbox compatibility	Data check
Temporal and spatial coverage and resolution	Scientific methodology	Archive	Expert evaluation
Providers	Uncertainty quantification		Dataset maturity
Dataset version	Validation		Key strengths and limitations
Data update	Inter-comparison		

Entries with the mark | display content that is specific for the variable selected



- “[...] **FAIR principles only covers part of best practice in RDM** and Open Research, e.g. data curation practices, data services, and data visualization.” (S.2.3 [fairsfair.gitbook.io](https://fairsfair.gitbook.io))
- “[...] **FAIR compliant data are not necessarily of high quality**, and the issue of quality assurance of the data is a separate one [...]” (S.2.2 [fairsfair.gitbook.io](https://fairsfair.gitbook.io))



# Anforderungen an Nachnutzbarkeit und Datenqualität

1. **“Fit for use”**:  
Gebrauchstaugliche Daten ([Chrisman, 1983](#)) —> FAIR-Prinzipien
2. **“Fit for purpose”**:  
Gütekriterien und weitere Parameter (u.a. Vollständigkeit) ([Chapman, 2005](#)) „**Datenqualität** hängt immer **vom Kontext** ab, in dem sie verwendet werden, ..... es gibt **keinen absoluten Qualitätsmaßstab**“



# CHATSTURM!!!

- Was sind (in deiner Fachdisziplin) relevante Qualitäts- oder Gütekriterien für Daten?

**Ziel: Beantworten einer (eigenen) Forschungsfrage**



## 2. Fit for Purpose

Drei Aspekte:

- Vertrauenswürdigkeit der Quelle/Daten  
Integrität

# Grundvoraussetzung: Bewertung der Integrität

integer?

- offizielle EU-Webseite
- Copernicus Programm der EU und ESA
- ECMWF internationales Forschungsinstitut (unabhängig)



The screenshot shows the header of the Copernicus website. It features the European Commission logo, the Copernicus logo with the tagline 'Europe's eyes on Earth', and the ECMWF logo with the text 'IMPLEMENTED BY'. To the right is the Climate Change Service logo. Below the logos is a navigation bar with links for Home, Search, Datasets, Applications, Toolbox, Support, and Live. The main content area below the navigation bar displays the text 'ERA5 hourly data on single levels from 1959 to present'.



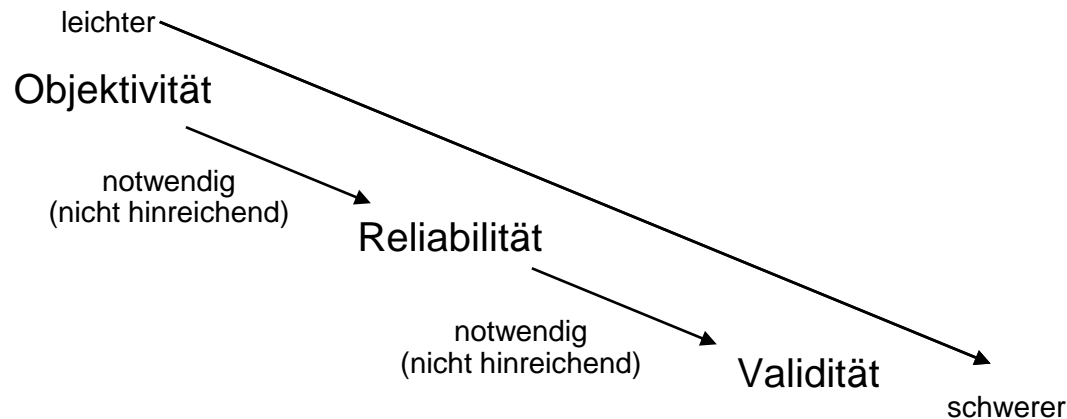
## 2. Fit for Purpose

### Drei Aspekte:

- Vertrauenswürdigkeit der Quelle/Daten  
Integrität
- klassische Gütekriterien  
Objektivität – Reliabilität – Validität

# Klassische Gütekriterien

- **Objektivität:** Unabhängigkeit von der Messenden Person (Ergebnis nicht durch messende Person beeinflusst)
- **Reliabilität:** Zuverlässigkeit des Messinstruments (gleiches Ergebnis bei gleichen Bedingungen)
- **Validität:** Gültigkeit des Messinstruments (Ergebnis spiegelt wieder, was gemessen werden sollte)





## 2. Fit for Purpose

### Drei Aspekte:

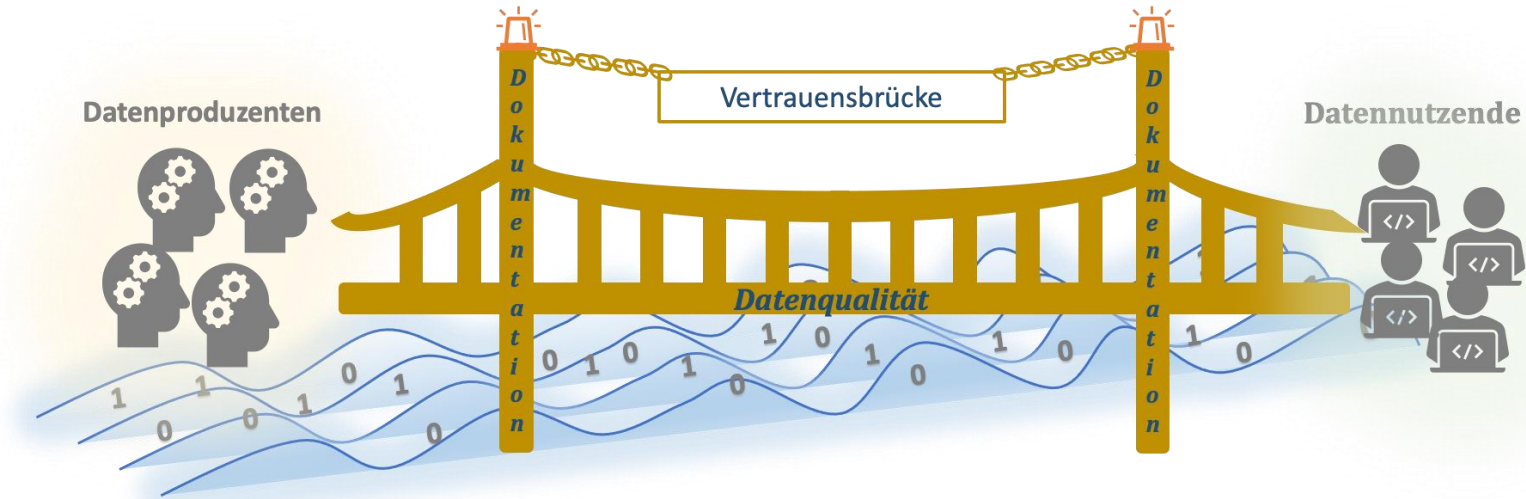
- Vertrauenswürdigkeit der Quelle/Daten  
Integrität
- klassische Gütekriterien  
Objektivität – Reliabilität – Validität
- Spezialfall: Reproduzierbarkeit/Replizierbarkeit  
Lassen sich Daten/Studie reproduzieren/replizieren?



- Forschungsvorhaben
- Kontext der Erhebung
- Erhebungsmethode
- Struktur der Daten und deren Beziehungen zueinander
- Qualitätsmaßnahmen
- Erklärungen für Codes und Labels (Codebook)
- Datenversionen und die enthaltenen Änderungen
- Informationen zum Zugang, Nutzungsbedingungen und Vertraulichkeit
- Dauerhafte Kontaktmöglichkeit



Quelle: [forschungsdaten.info \(2022\)](https://www.forschungsdaten.info/) und [Deutsche Forschungsgesellschaft \(2021\)](https://www.forschungsdaten.de/)



➤ *Datenqualität, die sich auf eine vollständige Datendokumentation stützt, baut eine Vertrauensbrücke zwischen Datenproduzenten und Datennutzende.*

Quelle für das „Bridge of Trust“-Konzept: [Sposito \(2017\)](#), [Ariza \(2023a\)](#)

# Wiederverwendbare

## Daten



*Fit for Purpose*

*Fit for use*

REUSABLE



DATENDOKUMENTATION

INTEROPERABLE



ACCESSIBLE



FINDABLE



FAIR - DATEN

DATENQUALITÄT

disziplinspezifische  
Qualitätsparameter

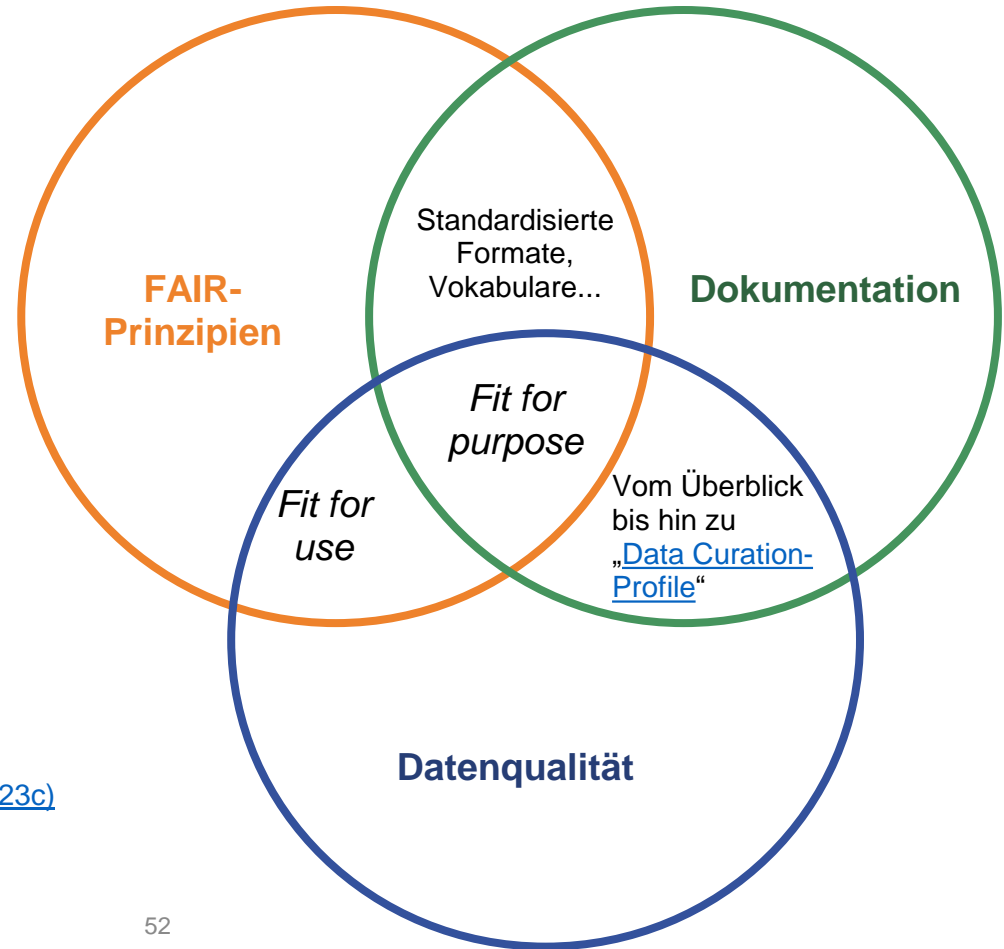
reproduzierbar  
valide  
zuverlässig  
objektiv



Quelle / Daten  
Integrität  
(vertrauenswürdig)

Beginn des Datenlebenszyklus

Wiederverwendbare  
Daten sind FAIR, gut  
dokumentiert und von  
guter Qualität!



Quelle für Data fitness: [Chapman \(2005\)](#), [Ariza \(2023c\)](#)



# Fragen und Antworten

# Vielen Dank für Eure Aufmerksamkeit

Angela Ariza

<https://orcid.org/0000-0002-1005-5780>

Esther Asef

<http://orcid.org/0000-0003-2411-4953>

Juliane Jacob

<https://orcid.org/0000-0002-0443-3570>

Andreas Mühlichen

<https://orcid.org/0000-0003-3115-4021>

Karsten Peters-von Gehlen

<https://orcid.org/0000-0003-0158-2957>

Hermann Schranzhofer

<https://orcid.org/0000-0003-0249-2726>

Ute Trautwein-Bruns

<https://orcid.org/0000-0003-0531-0182>



Alle Inhalte dieser Präsentation stehen unter der Lizenz [Creative Commons BY 4.0 International](https://creativecommons.org/licenses/by/4.0/).

Veröffentlicht unter <https://doi.org/10.5281/zenodo.7568266>

UAG [https://www.forschungsdaten.org/index.php/UAG\\_Schulungen/Fortbildungen](https://www.forschungsdaten.org/index.php/UAG_Schulungen/Fortbildungen)

# Anlage: Link-Sammlung und Quellen I

Ariza, Angela. (2023a). Data Quality - Bridge of Trust. <https://doi.org/10.6084/m9.figshare.21944333.v1>, zuletzt geprüft am 01.02.2023.

Ariza, Angela. (2023b). Reusable Data. <https://doi.org/10.6084/m9.figshare.21944393.v1>, zuletzt geprüft am 01.02.2023.

Ariza, Angela. (2023c). Data fitness. <https://doi.org/10.6084/m9.figshare.21944477.v1>, zuletzt geprüft am 01.02.2023.

Brettschneider, P., Biernacka, K., Böker, E., Danker, S. A., Jacob, J., Perry, A., Wiljes, C. & Wuttke, U. (2021). *Urheberrecht und Lizenzierung bei Forschungsdaten*. <https://doi.org/10.5281/ZENODO.5243232>

Chapman, A. (2005). *Principles of data quality*. [https://assets.ctfassets.net/uo17ejk9rkwj/2gupj7dJlw62UeOUYiqSsm/0a4bb732bd7fd8cf28f7703dc20a43ba/Data\\_Quality\\_-\\_ENGLISH.pdf](https://assets.ctfassets.net/uo17ejk9rkwj/2gupj7dJlw62UeOUYiqSsm/0a4bb732bd7fd8cf28f7703dc20a43ba/Data_Quality_-_ENGLISH.pdf), zuletzt geprüft am 01.02.2023.

Copernicus. *Climate reanalysis*. <https://climate.copernicus.eu/climate-reanalysis>, zuletzt geprüft am 01.02.2023.

Creative Commons. Creative Commons. <https://creativecommons.org/>, /, zuletzt geprüft am 01.02.2023.

Chrisman, N. R. (1983). The role of quality information in the long-term functioning of a geographic information system. *Cartographica The International Journal for Geographic Information and Geovisualization* (Vol. 1), <http://wellar.ca/wellarconsulting/auto-carto%20six%20proceedings%20-%20volume%201.pdf#page=315>, zuletzt geprüft am 01.02.2023.

DataCite. *DataCite Search*. <https://search.datacite.org/>, zuletzt geprüft am 01.02.2023.

Deutsche Forschungsgesellschaft. (2021, 21. Dezember). *Checkliste: Umgang mit Forschungsdaten*. [https://www.dfg.de/download/pdf/foerderung/grundlagen\\_dfg\\_foerderung/forschungsdaten/forschungsdaten\\_checkliste\\_de.pdf](https://www.dfg.de/download/pdf/foerderung/grundlagen_dfg_foerderung/forschungsdaten/forschungsdaten_checkliste_de.pdf), zuletzt geprüft am 01.02.2023.

## Link-Sammlung und Quellen II

Engelhardt, C. et al. (2022). *How to be FAIR with your data*. Göttingen University Press. <https://doi.org/10.17875/gup2022-1915>

EUDAT. *B2FIND*. <https://b2find.eudat.eu/>, zuletzt geprüft am 01.02.2023.

forschungsdaten.info. (2022, 14. November). *Datendokumentation*. <https://forschungsdaten.info/themen/beschreiben-und-dokumentieren/datendokumentation>, zuletzt geprüft am 01.02.2023.

Forschungsdaten.org. (2023, 11. Januar). *Mitglieder der UAG Schulungen/Fortbildungen*. [https://www.forschungsdaten.org/index.php/Mitglieder\\_der\\_UAG](https://www.forschungsdaten.org/index.php/Mitglieder_der_UAG), zuletzt geprüft am 01.02.2023.

Hassell, D., Gregory, J., Blower, J., Lawrence, B. N. & Taylor, K. E. (2017). A data model of the Climate and Forecast metadata conventions (CF-1.6) with a software implementation (cf-python v2.1). *Geoscientific Model Development*, 10(12), 4619–4646. <https://doi.org/10.5194/gmd-10-4619-2017>

Hersbach, H., Bell, B., Berrisford, P., Biavati, G., Horányi, A., Muñoz Sabater, J., Nicolas, J., Peubey, C., Radu, R., Rozum, I., Schepers, D., Simmons, A., Soci, C., Dee, D., Thépaut, J.-N. (2023, 1. Februar). *ERA5 hourly data on single levels from 1959 to present*. Copernicus Climate Change Service (C3S) Climate Data Store (CDS). <https://doi.org/10.24381/cds.adbb2d47>

Kadow, C., Wentzel, B., Jaekel, I., Bahlo, M. & Cubasch, U. (2017). Berlin Climate Record - Inner City - 1766-1934. <https://doi.org/10.1594/PANGAEA.870862>

Lacagnina, C., Doblás-Reyes, F., Larnicol, G., Buontempo, C., Obregón, A., Costa-Surós, M., San-Martín, D., Bretonnière, P.-A., Polade, S. D., Romanova, V., Putero, D., Serva, F., Llabrés-Brustenga, A., Pérez, A., Cavaliere, D., Membrive, O., Steger, C., Pérez-Zanón, N., Cristofanelli, P., . . . Díez, M. G. (2022). Quality Management Framework for Climate Datasets. *Data Science Journal*, 21(1), 10. <https://doi.org/10.5334/dsj-2022-010>



## Link-Sammlung und Quellen III

RCSB Protein Data Bank. *RCSB PDB: Policies*. <https://www.rcsb.org/pages/policies>, zuletzt geprüft am 01.02.2023.

re3data.org - Registry of Research Data Repositories. <https://doi.org/10.17616/R3D>

Redfern, N. (2021). A Data Set for US Horror Film Trailers. *Research Data Journal for the Humanities and Social Sciences*, 1–12. <https://doi.org/10.1163/24523666-bja10017>

Redfern, N. (2021). *US\_Horror\_Trailers\_Data\_Set*. <https://doi.org/10.5281/zenodo.4479068>

*Research Data Journal for the Humanities and Social Sciences*. <https://brill.com/view/journals/rdj/rdj-overview.xml>, zuletzt geprüft am 01.02.2023

SowiDataNet|datorium. *Nutzungsbedingungen*. <https://data.gesis.org/sharing/#!TermsOfUse>, zuletzt geprüft am 01.02.2023.

Sposito, F. A. (2017). What do data curators care about? Data quality, user trust, and the data reuse plan. <https://library.ifla.org/id/eprint/1797/1/S06-2017-sposito-en.pdf>, zuletzt geprüft am 01.02.2023.

Universitätsbibliothek Regensburg. *DBIS: Fachübersicht*. <https://dbis.ur.de/fachliste.php>, zuletzt geprüft am 01.02.2023.

Wilbrandt, J. (2021). *Daten (nach)nutzen — FAIR, aber wie? [Workshop]*. Zenodo. <https://doi.org/10.5281/zenodo.5554816>