



Workshop: Forschungsdatenmanagement in den Sprach- und Informationswissenschaften - warum und wie?

Mittwoch, 25. September 2019 um 10:00 Uhr
(Annette Strauch, M.A.)

Workshop: Forschungsdatenmanagement in den Sprach- und Informationswissenschaften - warum und wie?

Love your data!

- Forschungsdatenmanagement – warum und wie?
- Grundlagen des Forschungsdatenmanagement für die Sprach- und Informationswissenschaften
- Essentials der Erstellung eines Datenmanagementplans (z.B. mit RDMO)

Universitätsbibliothek Hildesheim: Workshop: Forschungsdatenmanagement in den Sprach- und Informationswissenschaften - warum und wie?

& Finden Forschen & Publizieren Schulung & Beratung Aktuelles & Veranstaltungen Über uns

The screenshot shows a navigation menu on the left with the following items: 'Forschungsdatenmanagement', 'Publizieren', 'Universitätsverlag', 'HilData', 'HilDok', 'DOI für Open Access-Publikationen auf HilDok', 'Open Access', and 'Open-Access-Publikationsfonds'. The main content area on the right is titled 'Ansprechpartner' and features the 'forschungsdaten.info' logo. Below the logo, it lists 'Forschungsdatenmanagement' with contact details for Annette Strauch, M. A., including her title, telephone number (+49 5121 883-93074), and a 'Mail' link. At the bottom of this section is the 'Universitätsverlag' logo and contact information for PD Dr. Mario Müller, Editorial Director Print & Digital Publishing, with a telephone number (+49 5121 883-02084).

<https://www.uni-hildesheim.de/bibliothek/forschen-publizieren/forschungsdatenmanagement/>

Universitätsbibliothek Hildesheim:

**Workshop: Forschungsdatenmanagement in den Sprach-
und Informationswissenschaften - warum und wie?**



Universitätsbibliothek Hildesheim: Workshop: Forschungsdatenmanagement in den Sprach- und Informationswissenschaften - warum und wie?

Wer bin ich?

- Ansprechpartnerin für das
Forschungsdatenmanagement an der SUH

Annette Strauch, M.A. (Anglistik, Europäische Ethnologie)

Stiftung Universität Hildesheim
Universitätsplatz 1
31141 Hildesheim

annette.strauch@uni-hildesheim.de

Telefon: +49 (0) 51 21 – 883 93074,

<https://www.uni-hildesheim.de/forschungsdaten/>

Universitätsbibliothek Hildesheim:

Workshop: Forschungsdatenmanagement in den Sprach- und Informationswissenschaften - warum und wie?

Workshop mit Teilnehmerinnen und Teilnehmern aus folgenden Instituten:

- Institut für Sprach- und Informationswissenschaften
- Institut für Informationswissenschaften und Sprachtechnologie
- Institut für Erziehungs- und Sozialwissenschaften (Politikwissenschaften)
- Institut für Übersetzungswissenschaften und Fachkommunikation
- Institut für deutsche Sprache und Literatur

Universitätsbibliothek Hildesheim:

Workshop: Forschungsdatenmanagement in den Sprach- und Informationswissenschaften - warum und wie?

Forschungsinteressen:

- Text(sorten)linguistik, Diskurslinguistik, Metaphernforschung, Sprachideologien, Sprachliche Normen und Varianz, Literaturdidaktik, Literaturwissenschaft
- Autokratieforschung, Migrationspolitik, Politik und Internet
- Computerlinguistik

Universitätsbibliothek Hildesheim:

Workshop: Forschungsdatenmanagement in den Sprach- und Informationswissenschaften - warum und wie?

Fragen und Bedarfe an das Forschungsdatenmanagement für diesem Workshop:

- Wunsch nach einer generellen Einführung
- Datenmanagementplan (was muss hinein?)
- Einwilligungen, rechtl. Themen
- Infrastrukturen

Universitätsbibliothek Hildesheim: Workshop: Forschungsdatenmanagement in den Sprach- und Informationswissenschaften - warum und wie?

Forschungsdatenmanagement (FDM) (...)

(...) bezeichnet den systematischen und planvollen Umgang mit Forschungsdaten über den gesamten Lebenszyklus der Daten hinweg.



(Photo by [rawpixel](#) on [Unsplash](#))



Workshop: Forschungsdatenmanagement in den Sprach- und Informationswissenschaften - warum und wie?

“Research data management is an explicit process covering the creation and stewardship of research materials to enable their use for as long as they retain value.”

(Whyte, A. and Rans, J., [Glossary Digital Curation Center](#))

"Das Forschungsdatenmanagement muss so gestaltet werden, dass Datenzugriff und -auswertung unabhängig vom Datenerzeuger möglich wird und bleibt."

(Büttner, Hobohm und Müller 2011 S. 14, [Referenz](#))

**Workshop: Forschungsdatenmanagement in den Sprach-
und Informationswissenschaften - warum und wie?**

*“Linguistic data arise from
a range of techniques
that straddle traditional humanities
and scientific methodologies, from field-
based observation to corpus data to
experimentation.”*

'Linguistic data management' by N. Thieberger and A.L. Berez,
in: The Oxford Handbook of Linguistic Fieldwork (2012).

Workshop: Forschungsdatenmanagement in den Sprach- und Informationswissenschaften - warum und wie?

Die Reflexion über Forschungsdaten

- nicht nur wissenschaftsübergreifend
- sondern aus der Perspektive einzelner Disziplinen

Universitätsbibliothek Hildesheim:

Workshop: Forschungsdatenmanagement in den Sprach- und Informationswissenschaften - warum und wie?

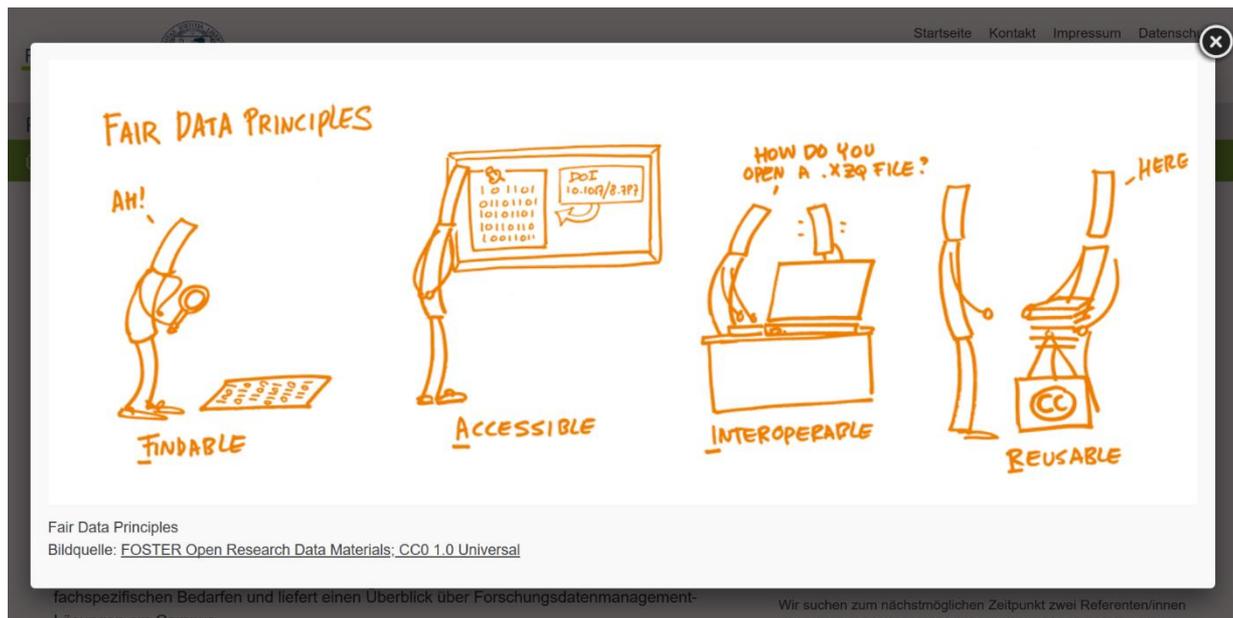
Forschungsdaten

- Text- und Audiokorpora
- elektronische Textausgaben und –sammlungen
- Gesprächstranskriptionen
- Interviewdaten

Universitätsbibliothek Hildesheim: Workshop: Forschungsdatenmanagement in den Sprach- und Informationswissenschaften - warum und wie?

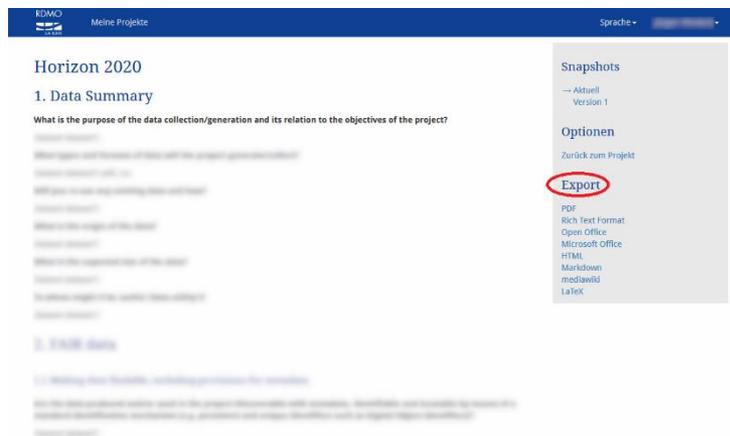
FAIR Principles.

Metadaten und Daten sollen auffindbar, zugänglich,
interoperabel und wiederverwendbar sein (Wilkinson u.a.
2016, [Referenz](#)).



Universitätsbibliothek Hildesheim: Workshop: Forschungsdatenmanagement in den Sprach- und Informationswissenschaften - warum und wie?

Die Forschungsförderer verlangen zunehmend ein transparentes, auf Öffentlichkeit ausgerichtetes, Forschungsdatenmanagement.



Ein Datenmanagementplan strukturiert den Umgang mit den Forschungsdaten (z.B. mit RDMO).

Universitätsbibliothek Hildesheim: Workshop: Forschungsdatenmanagement in den Sprach- und Informationswissenschaften - warum und wie?



Willkommen bei RDMO Hildesheim

Das Ziel des RDMO Projekts ist es, eine Webapplikation bereitzustellen, die die strukturierte Planung, Umsetzung und Verwaltung der Daten in einem wissenschaftlichen Projekt unterstützt. Zusätzlich sollen die gesammelten Informationsdaten in textueller Form für Anforderungen von Referenzen oder für Berichte ausgabefähig sein.

Diese Webseite ist ein Prototyp zur Veranschaulichung der bereits implementierten Funktionen.

Wenn Sie mehr über das Projekt erfahren wollen, besuchen Sie rdmorganiser.github.io.

Universitätsbibliothek Hildesheim:

Workshop: Forschungsdatenmanagement in den Sprach- und Informationswissenschaften - warum und wie?

Ein DMP beschreibt, wie während der Laufzeit und nach Projektende mit verwendeten Daten verfahren wird.

Projektbeschreibung:

Allgemeine Angaben zum Projektvorhaben, wie z.B. Ziele, Projektträger und Laufzeit.

Universitätsbibliothek Hildesheim:

Workshop: Forschungsdatenmanagement in den Sprach- und Informationswissenschaften - warum und wie?

Existierende Arten von Daten:

Beschreibung der existierenden Daten, die für das Projekt nachgenutzt werden können und ihre Integration in das Projektvorhaben.

Universitätsbibliothek Hildesheim:

Workshop: Forschungsdatenmanagement in den Sprach- und Informationswissenschaften - warum und wie?

Datenorganisation:

Angaben zur einheitlichen Regelung zur Datenverwaltung innerhalb des Projektes (z.B. bezüglich Dateispeicherung, Dateibenennung, Synchronisation, Versionierung und Kollaborativen Arbeiten).

Weitere Informationen hier:

<https://zenodo.org/record/3234414#.XYoSInvgriU>

Universitätsbibliothek Hildesheim: Workshop: Forschungsdatenmanagement in den Sprach- und Informationswissenschaften - warum und wie?

Beispiele für DMP-Werkzeuge für Projekte für die Sprach-
und Informationswissenschaften.



Universitätsbibliothek Hildesheim: Workshop: Forschungsdatenmanagement in den Sprach- und Informationswissenschaften - warum und wie?

CLARIN-D (Common Language Resources and Technology Infrastructure) Wizard

Wizard zur Planerstellung für das
Datenmanagement (experimentell)

- 1 PROJEKTINFORMATIONEN
- 2 FORSCHUNGSDATEN
- 3 DOKUMENTATION UND METADATEN
- 4 ETHIK UND RECHT
- 5 DATENSPEICHERUNG UND BACKUP
- 6 DATENAUSWAHL - ORGANISATION UND -ARCHIVIERUNG
- 7 DATEN-SHARING
- 8 ZEITLICHER RAHMEN
- 9 BUDGET

Administrative Daten zum Projekt

Welches Forschungsprojekt wird die Daten bereitstellen?

Was sind die Ziele des Projektes?

Wer ist verantwortlicher Projektleiter?

Wie ist der Name der Forschungseinrichtung bzw. Universität?

<https://www.clarin-d.net/de/aufbereiten/datenmanagementplan-entwickeln>

Universitätsbibliothek Hildesheim: Workshop: Forschungsdatenmanagement in den Sprach- und Informationswissenschaften - warum und wie?

CLARIN-D und DARIAH-DE sind spezialisierte
Forschungsinfrastrukturen.

Sie stimmen sich eng im Bereich der technischen
Infrastruktur ab, bei Basisdiensten, Metadaten,
gemeinsamen Komponenten, etc.



Universitätsbibliothek Hildesheim: Workshop: Forschungsdatenmanagement in den Sprach- und Informationswissenschaften - warum und wie?



Universitätsbibliothek Hildesheim: Workshop: Forschungsdatenmanagement in den Sprach- und Informationswissenschaften - warum und wie?

The screenshot shows the CLARIN-D website homepage. At the top left, there are flags for Germany and the UK. The CLARIN-D logo, featuring a stylized molecular structure, is prominently displayed. To the right of the logo is a search bar with the placeholder text "Suchen ...". Further right are social media icons for Twitter, YouTube, and Facebook. Below the logo and search bar is a dark red navigation bar with the following menu items: Home, Auffinden, Auswerten, Aufbereiten, Facharbeitsgruppen, Blog, Über..., Hilfe, and Aktuelles. Below the navigation bar, the word "WILLKOMMEN" is written in bold. Underneath, there are three image-based tiles. The first tile shows a hand hovering over a search bar with the text "Suchen und finden". The second tile shows a person in a suit pointing at a cloud with arrows, symbolizing cloud computing or data management. The third tile shows a row of metal filing cabinets, with one cabinet open, symbolizing data storage or archiving.

Universitätsbibliothek Hildesheim: Workshop: Forschungsdatenmanagement in den Sprach- und Informationswissenschaften - warum und wie?



The screenshot shows the CLARIN-D website interface. At the top left, there are flags for Germany and the UK. The CLARIN-D logo is prominently displayed. A search bar with the text 'Suchen ...' is located in the top right. Below the logo, a dark red navigation bar contains the following menu items: Home, Auffinden, Auswerten, Aufbereiten, Facharbeitsgruppen, Blog, Über..., Hilfe, and Aktuelles. The main content area features the heading 'CLARIN-D Legal Helpdesk' and a paragraph stating: 'Der CLARIN-D Legal Helpdesk bietet Ihnen folgende allgemeine Informationen bezüglich juristischer Fragestellungen in den Geisteswissenschaften:'. Below this, there are three columns of content:

- Legal Information Platform**
Beachten Sie: Ein Teil der Legal Information
- CLARIN Lizenzen**
Sprachressourcen werden – bedingt durch den jeweiligen individuellen Lizenzumfang – den folgenden CLARIN
- Publikationen**
Der Inhalt der Legal Information Platform wird nach und nach weiterentwickelt, bis er als Publikation veröffentlicht wird

Universitätsbibliothek Hildesheim:

Workshop: Forschungsdatenmanagement in den Sprach- und Informationswissenschaften - warum und wie?

Besondere Services und Angebote von CLARIN-D:

- Tool zum Erstellen eines Datenmanagementplans
- Referenzdatensätze, z.B. Deutsches Textarchiv (DTA)
- Deutsches Referenzkorpus (DeReKo)
- GermaNet, Digitales Wörterbuch der deutschen Sprache (DWDS)
- Deutsche Parlamentsdebatten (GermaParl)
- zahlreiche webbasierte Analysewerkzeuge für sprachbasierte Daten

Universitätsbibliothek Hildesheim: Workshop: Forschungsdatenmanagement in den Sprach- und Informationswissenschaften - warum und wie?



Suchen ...

Home [Auffinden](#) ▾ [Auswerten](#) ▾ [Aufbereiten](#) ▾ [Facharbeitsgruppen](#) ▾ [Blog](#) [Über...](#) ▾ [Hilfe](#) ▾ [Aktuelles](#)

VLO: Suchmaschine zum Auffinden von Forschungsdaten

Das VLO (Virtual Language Observatory) ist eine spezielle Suchmaschine zum Auffinden von Forschungsdaten im Umfeld von CLARIN. Es besteht im Wesentlichen aus einem großen Katalog an Forschungsdaten, der auf verschiedene Weise durchsucht werden kann. Über eine Textsuche kann nach Schlüsselbegriffen aus den Beschreibungen und nach Titeln von Ressourcen gesucht werden. Mit einer facettierten Suche kann die Suche aber auch eingeschränkt werden, um etwa gezielt nach speziellen Sprachen, Genres, etc. zu suchen.

Search

Search in the [Virtual Language Observatory](#) for language resources in the CLARIN repositories:

(e.g. search for a language or a type of resource)

Search

Universitätsbibliothek Hildesheim:

Workshop: Forschungsdatenmanagement in den Sprach- und Informationswissenschaften - warum und wie?

Der CLARIN-D User Guide ist ein praktischer Leitfaden für die Anpassung und Integration existierender Sprachressourcen in die CLARIN-D-Infrastruktur.

Universitätsbibliothek Hildesheim: Workshop: Forschungsdatenmanagement in den Sprach- und Informationswissenschaften - warum und wie?

Metadaten

INTERNATIONAL
STANDARD

ISO
24622-2

First edition

Language resource management —
Component metadata infrastructure
(CMDI) —

Part 2:
Component metadata specification
language

Gestion des ressources en linguistique — Coopération infrastructure de
métadonnées (CMDI) —

Partie 2: Composante métadonnées spécifique en linguistique

Universitätsbibliothek Hildesheim: Workshop: Forschungsdatenmanagement in den Sprach- und Informationswissenschaften - warum und wie?

Allg. Sprachwissenschaft &
Computerlinguistik

Forschung

Projekte

Abgeschlossene Projekte

Ressourcen

CLARIN Center

Repository

Gehostete Ressourcen

Mitarbeiter

Dissertationen

Habilitationen

Kontakt

CLARIN Center Tübingen

CLARIN-D is developing a digital infrastructure for language-centered research in the social sciences and humanities. The main function of the service centers in CLARIN-D is to provide relevant, useful data and tools in an integrated, interoperable and scalable way. The CLARIN-D infrastructure is designed in close collaboration with expert scholars in the humanities and social sciences, to ensure that it meets the needs of users in a systematic and easily accessible way.

CLARIN-D is built on the achievements of the preparatory phase of the European CLARIN initiative as well as CLARIN-D's Germany-specific predecessor project D-SPIN.

Tübingen, as coordinator of the CLARIN-D project, participates in nearly all aspects of the project.



This is a certified Type B CLARIN Center.



Universitätsbibliothek Hildesheim: Workshop: Forschungsdatenmanagement in den Sprach- und Informationswissenschaften - warum und wie?

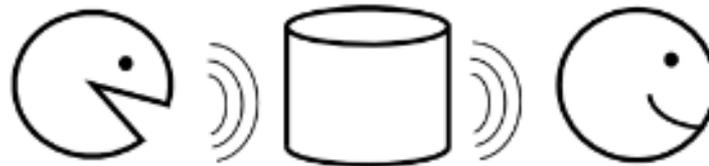
CLARIN-D Repositorium (Eberhard Karls Universität Tübingen)

Allg. Sprachwissenschaft & Computerlinguistik	Repository
Forschung	The CLARIN repository at the University of Tübingen offers long-term preservation of digital resources, along with their descriptive metadata.
Projekte	The mission of the repository is to ensure the availability and long-term preservation of resources, to preserve knowledge gained in research, to aid the transfer of knowledge into new contexts, and to integrate new methods and resources into university curricula.
Abgeschlossene Projekte	
Ressourcen	The repository is part of the eScience infrastructure of the University of Tübingen, which is a core facility that strongly cooperates with the library and computing center of the university.
CLARIN Center	
Repository	Integration of the repository into the national CLARIN-D and international CLARIN infrastructures gives it wide exposure, increasing the likelihood that the resources will be used and further developed beyond the lifetime of the projects in which they were developed.
Access	
Depositing	
Agreements and Guidelines	Among the resources currently available in the Tübingen Center Repository, researchers can find widely used treebanks of German (e.g. TüBa-D/Z), the German wordnet (GermaNet), the first manually annotated digital treebank (Index Thomisticus), as well as descriptions of the tools used by the WebLicht ecosystem for natural language processing.
Privacy Policy	
Preservation Policy	
Gehostete Ressourcen	 This repository was awarded the Core Trust Seal in March, 2019.
Mitarbeiter	
Dissertationen	
Habilitationen	 This repository was awarded the Data Seal of Approval in 2013 and 2015.
Kontakt	



Universitätsbibliothek Hildesheim:

**Workshop: Forschungsdatenmanagement in den Sprach-
und Informationswissenschaften - warum und wie?**



Bavarian Archive for Speech Signals

Universitätsbibliothek Hildesheim:

Workshop: Forschungsdatenmanagement in den Sprach- und Informationswissenschaften - warum und wie?

Werkzeuge und Dienste von CLARIN-D

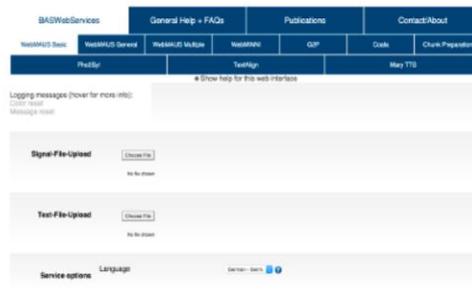
<https://www.clarin-d.net/de/sprachressourcen-und-dienste/werkzeuge-und-dienste>

Fragen an: corpora@uni-hamburg.de

Universitätsbibliothek Hildesheim: Workshop: Forschungsdatenmanagement in den Sprach- und Informationswissenschaften - warum und wie?

Beispiele für die Verwendung der CLARIN-D Infrastruktur

WebMAUS-Basic: Automatisches phonetisches Labelling und Segmentieren einer einzelnen deutschen Aufnahme mit Text



Interviews und Gespräche werden häufig aufgenommen und anschließend verschriftlicht, also transkribiert. Der Werkzeugsatz WebMAUS, der in der CLARIN-Infrastruktur zur Verfügung steht, erlaubt es, die Audioaufnahmen und Transkriptionen miteinander zu verknüpfen.

Besonders interessant für

WebMAUS Pipeline: Chunk-Vorbereitung langer Video-Interviews



Es ist eine Herausforderung mehrstündige Videoaufnahmen zeitlich zu alignieren. Hierfür bietet das BAS einen Webservice an, welcher die Aufnahmen automatisch in so genannte Chunks aufteilt, und diese wiederum in einem gemeinsamen File speichert. Dies wird im Folgenden Anwendungsszenario demonstriert.

Besonders interessant für Forschende der:

- Linguisten

<https://www.clarin-d.net/de/hilfe/beispielverwendungen-fuer-clarin>

Universitätsbibliothek Hildesheim: Workshop: Forschungsdatenmanagement in den Sprach- und Informationswissenschaften - warum und wie?

DGD
DATENBANK FÜR
GESPROCHENES
DEUTSCH

REGISTRIERUNG
LOGIN

E-Mail-Adresse

ANMELDEN

PASSWORT ÄNDERN
PASSWORT VERGESSEN
REGISTRIERUNG LOSCHEN

HERZLICH WILLKOMMEN IN DER DATENBANK FÜR GESPROCHENES DEUTSCH (DGD)

Die DGD wird für Forschung und Lehre bereitgestellt. Das System ermöglicht Ihnen den Zugriff auf ausgewählte Korpora des [Archivs für Gesprochenes Deutsch \(AGD\)](#), u. a. auf:

- das "Forschungs und Lehrkorpus gesprochenes Deutsch" (FOLK)
- das Korpus "Deutsche Mundarten" (Zwiemer-Korpus)
- das Korpus "Deutsche Umgangssprachen" (Pfeffer-Korpus)
- das Korpus "Gesprochene Wissenschaftssprache Kontrastiv" (GeWis, GWSS)
- das Korpus "Deutsch Heute" (DH)

Aus rechtlichen Gründen bitten wir Sie um eine einmalige kostenlose Registrierung. Vor Ihrer Registrierung können Sie sich anhand des Menüpunktes "Über die DGD" über das Angebot der DGD informieren.

Für Hinweise auf mögliche Probleme bei der Nutzung wären wir sehr dankbar.

Unsere Kontaktadresse lautet: dgd@ids-mannheim.de

english version

QUASSELSTRIPPEN
UND HOBBYFILMER
DAS FOLK
BRÄUCHT
DICH!

IDS
LEIBNIZ-INSTITUT FÜR
DEUTSCHE SPRACHE

Aktuelles | Organisation | Forschung | Onlineangebote | Bibliothek | Service | Publikationen | Über uns

Zentrale Forschung | Grammatik | Lexik | Pragmatik | **Digitale Sprachwissenschaft** | Öffentlichkeitsarbeit | Informationstechnik | Verwaltung

Startseite :: Organisation :: Direktion :: Korpuslinguistik/KorporaLinguistik :: Projekte :: Korpusausbau

Digitale Sprachwissenschaft

- :: Korpuslinguistik
- :: Aktuelles
- :: Projekte
- :: **Korpusausbau**
 - :: Einsatzgebiete
 - :: Akquisition
 - :: IDS-Textmodell
 - :: Archive
 - :: Ausgaben
 - :: Verfügbarkeit
 - :: Annotationen
 - :: Textgeber
 - :: Ehemalige Mitarbeiter
 - :: Kooperationen
 - :: Bibliographie (IDS)

Ausbau und Pflege der Korpora geschriebener Gegenwartssprache

DAS DEUTSCHE REFERENZKORPUS – DEREKO

die weltweit größte Sammlung deutschsprachiger Korpora als empirische Basis für die linguistische Forschung

Die Korpora geschriebener Gegenwartssprache des IDS

- bilden mit 43 Milliarden Wörtern (Stand 18.03.2019) die weltweit größte linguistisch motivierte Sammlung elektronischer Korpora mit geschriebenen deutschsprachigen Texten aus der Gegenwart und der neueren Vergangenheit.
- sind über COSMAS II kostenlos abfragbar
- enthalten belletristische, wissenschaftliche und populärwissenschaftliche Texte, eine große Zahl von Zeitungstexten sowie eine breite Palette weiterer Textarten und werden kontinuierlich weiterentwickelt.

Kontakt:
<korpuslinguistik@ids-...>

Leitung:
Dr. Marc Kupietz
<kupietz@ids-...>

Wissenschaftliche Mitarbeiter:
Cyril Belica <belica@ids-...>
Dr. Harald Lungen
<luengen@ids-...>
Rainer Perkuhn
<perkuhn@ids-...>

Kooperationen:
siehe hier

Ehemalige am Korpusaufbau

Universitätsbibliothek Hildesheim: Workshop: Forschungsdatenmanagement in den Sprach- und Informationswissenschaften - warum und wie?



IDS Repository

[Description](#)
[Terms of Use](#)
[Browse](#)
[Search](#)
[About](#)
[OAIprovider](#)



certification logo

Welcome to the IDS Repository

Welcome to the research data repository at the [Leibniz-Institut für Deutsche Sprache](#) (IDS Mannheim).

The [IDS](#) Repository aims at long-term archival of linguistic resources and tools in the field of German studies. It provides data together with metadata in [Dublin Core](#) and [CMD1](#) formats.

The [IDS](#) is member of [CLARIN](#) and the national [CLARIN-D](#).

The [Virtual Language Observatory](#) harvests our metadata and provides a faceted search interface for them.



News

- 18 March 2019**
The IDS Repository has been recertified as a [CLARIN B Center](#).
- 18 March 2019**
The [IDS](#) repository has now been certified with the [Core Trust Seal](#).
- 11 March 2019**
presentation of the [IDS](#) repository at [IDS](#).
- 22 February 2019**
presentation of the [IDS](#) repository at the

Universitätsbibliothek Hildesheim: Workshop: Forschungsdatenmanagement in den Sprach- und Informationswissenschaften - warum und wie?

Anmelden (DTAQ) DWDR

DTA Werke im Deutschen Textarchiv

Texte Projekt Dokumentation

suchen

In den Titeldaten Im Korpus In der Dokumentation Hilfe

Beispielanfragen: ehelichen with \$p=VVNF *zeit with \$p=NN ehelich with \$p=ADJA

Deutsches Textarchiv

GRUNDLAGE FÜR EIN REFERENZKORPUS DER NEUHOCHDEUTSCHEN SPRACHE

Das Deutsche Textarchiv stellt einen disziplinen- und gattungsübergreifenden Grundbestand deutschsprachiger Texte aus dem Zeitraum von ca. 1600 bis 1900 bereit. Die Textauswahl erfolgte auf der Grundlage einer von Akademikern erstellten und ausführlich kommentierten, umfangreichen Bibliographie. In Ergänzung wurden einschlägige Literaturgeschichten und (Fach-)Bibliographien ausgewertet. Aus der Gesamtliste der auf diesem Wege ermittelten Titel wurde von der DTA-Projektgruppe ein hinsichtlich der repräsentierten Textsorten und Disziplinen ausgewogenes Korpus zusammengestellt (weitere Informationen zur Textauswahl).

Um den historischen Sprachstand möglichst genau abzubilden, werden als Vorlage für die Digitalisierung in der Regel die Erstausgaben der Werke zugrunde gelegt. Das elektronische Volltextkorpus des DTA ist über das Internet frei zugänglich und dank seiner Aufbereitung

Das DTA in Zahlen

Sie können in 3 960 Werken recherchieren.
Weitere Zahlen finden Sie hier.

Das DTA am 21. September 2019

Am 21.09.1939 w
ehemaligen Bauer
Winzerdorf Coswig
Sachsen das Stadt
verliehen.

Führer durch Coswig, Kötl
Coswig und Umgegend,
Kötzschenbroda-Dresden,

Neue Werke im DTA

U+H Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

h_zsk hamburg center
für sprachkorpora

SK KORPORA WERKZEUGE SUPPORT

Login

English Deutsch

Das Hamburger Zentrum für Sprachkorpora (HZSK)

<u>Linguistische Korpora</u> Aufbau und Pflege des HZSK Repository für die nachhaltige Verfügbarmachung linguistischer Daten und Dienste	<u>Korpus-Workflows</u> Beratung von Forschern und Forscherinnen vor und während der Erstellung und Analyse von digitalen Sprachkorpora	<u>Korpusaufbereitung</u> Mitwirkung bei der Aufbereitung von fertig gestellten Sprachkorpora zur Veröffentlichung, Archivierung und Weiterverwendung	<u>Werkzeuge</u> Entwicklung von Werkzeugen zur Erstellung, Verwaltung, Analyse, Archivierung und Veröffentlichung von Sprachkorpora	<u>Best Practices</u> Entwicklung, Dokumentation und Vermittlung von Best Practices und Standards für die Erstellung und Vorhaltung digitaler Sprachkorpora
--	--	--	---	--

Das HZSK schafft einen institutionellen Rahmen für die nachhaltige Nutzbarkeit von linguistischen Forschungsprimärdaten. Als Zusammenschluss von Angehörigen verschiedener Fachbereiche und Einrichtungen der Universität Hamburg besteht das HZSK über den Rahmen von zeitlich befristeten Projekten hinaus und dient somit der Förderung und Koordination computergestützter empirischer Forschung und Lehre in den Sprachwissenschaften und angrenzenden Disziplinen an der Universität Hamburg.

Universitätsbibliothek Hildesheim: Workshop: Forschungsdatenmanagement in den Sprach- und Informationswissenschaften - warum und wie?



The screenshot displays the Linguistic Data Consortium (LDC) website. At the top, the LDC logo is shown with the text "Linguistic Data Consortium". Below this, the title "THE WORLD ATLAS OF LANGUAGE STRUCTURES ONLINE" is displayed above a world map. A navigation menu includes "Home", "Features", "Chapters", "Languages", "References", and "Authors". A secondary menu contains "Changes", "Credits", "Legal", "Download", and "Contact". The main content area features a "Welcome to WALS Online" section, a search bar, and a "Latest Comments" section with two entries.

Linguistic Data Consortium

THE WORLD ATLAS OF LANGUAGE STRUCTURES ONLINE

Home Features Chapters Languages References Authors

Changes Credits Legal Download Contact

Welcome to WALS Online

The World Atlas of Language Structures (WALS) is a large database of structural (phonological, grammatical, lexical) properties of languages gathered from descriptive materials (such as reference grammars) by a team of 55 authors.

The first version of WALS was published as a book with CD-ROM in 2005 by Oxford University Press. The first online version was published in April 2008. The second online version was published in April 2011.

The 2013 edition of WALS corrects a number of coding errors especially in Chapters 1 and 3. A full list of changes is available here.

Starting with this edition of WALS, there will not be specific editions every two or three years, but we will be updating it whenever corrections or additions are made. Changes in value assignment will be made transparent by showing a history on the respective pages.

Google Custom Search

Latest Comments

Comment on Datapoint for feature 49A and language wals_code_scr by Rene
Wed Sep 18 2019
For all intents and purposes, modern standard BCS has 6 morphological ...

Comment on Datapoint Wathawurrung / Position of Case Affixes by Corrina Eccles
Mon Aug 05 2019

Universitätsbibliothek Hildesheim: Workshop: Forschungsdatenmanagement in den Sprach- und Informationswissenschaften - warum und wie?

Werkzeuge und Dienste von CLARIN-D



The screenshot shows the CLARIN-D website with a search bar at the top right. The main navigation bar includes links for Home, Auffinden, Auswerten, Aufbereiten, Facharbeitsgruppen, Blog, Über..., Hilfe, and Aktuelles. The main content area is titled "Unterstützung von Enhanced Publications in CLARIN: Zitation, Archivierung und Zugang zu Forschungsdaten". It features a section for "BAC CLARIN Repository" with a list of publications and a "Corpus PID" section. The text explains the importance of research data and provides a solution for accessing it.

Unterstützung von Enhanced Publications in CLARIN: Zitation, Archivierung und Zugang zu Forschungsdaten

Besonders interessant für Forschende aus den

- Geisteswissenschaften, die mit empirischen Sprachdaten arbeiten möchten
- Sprachtechnologien

Ausgangslage:

In einer Publikation sind Daten referenziert mittels PID eines Korpus' gesprochener Sprache PD2: 11858/00-1779-0000-001F-88A9-6

Ziel:

Verifikation des Forschungsergebnisses oder Nachnutzung der Daten durch eigene Untersuchung des zitierten Korpus' gesprochener Sprache

Lösung:

Die Daten zu einem Identifikator werden in eine Repository lokalisiert, man meldet sich an und lädt das Korpus herunter

Verwandte CLARIN-D-Werkzeuge und -Dienste

- Vergleichende Textanalyse auf Wortebene
- Zugang zu einer Auswahl von Sprachaufnahmen und -annotationen per autorisiertem Download

<https://www.clarin-d.net/de/herunterladen-eines-bas-korpus>

Universitätsbibliothek Hildesheim: Workshop: Forschungsdatenmanagement in den Sprach- und Informationswissenschaften - warum und wie?



Digitale Forschungsinfrastruktur für die
Geistes- und Kulturwissenschaften



EINSTIEG



FORSCHUNGSDATEN



DIENSTE UND
WERKZEUGE



VERANSTALTUNGEN



BERATUNG



TEXTGRID

DARIAH-DE unterstützt die mit digitalen Ressourcen und Methoden arbeitenden Geistes- und KulturwissenschaftlerInnen in Forschung und Lehre. Dazu baut der Verbund eine digitale Forschungsinfrastruktur auf und entwickelt Materialien für Lehre und Weiterbildung im Bereich der Digital Humanities (DH).

Der Beginn der Betriebskooperation ab März 2019 markiert eine neue Phase in der Entwicklung von DARIAH-DE. 16 Partnerinstitutionen haben durch den Abschluss einer Betriebskooperationsvereinbarung den Grundstein für den nachhaltigen Betrieb und somit für die Verstetigung von DARIAH-DE gelegt. Im Rahmen dessen werden wir unseren Austausch mit weiteren Initiativen der digitalen Forschung und Lehre in den Geistes- und Kulturwissenschaften intensivieren. Insbesondere eine noch stärkere Kooperation mit CLARIN-D soll dazu beitragen, unsere Vision eines kooperativen Großraums zur digitalen Forschung in den europäischen Geistes- und Kulturwissenschaften weiter voranzubringen. Zu diesem Zweck schließen sich CLARIN-D und DARIAH-DE ab 2019 zu der gemeinsamen Infrastruktur CLARIAH-DE zusammen, die die Dienste beider Initiativen bündeln wird.

Universitätsbibliothek Hildesheim:

Workshop: Forschungsdatenmanagement in den Sprach- und Informationswissenschaften - warum und wie?

Urheberrecht in der Wissenschaft

https://www.bmbf.de/upload_filestore/pub/Handreichung_UrhWissG.pdf

Rechtliche Themen im FDM, siehe: Gutachten, DataJus, https://tu-dresden.de/gsw/jura/igetem/jfbimd13/ressourcen/dateien/publikationen/DataJus_Zusammenfassung_Gutachten_12-07-18.pdf?lang=de

Universitätsbibliothek Hildesheim: Workshop: Forschungsdatenmanagement in den Sprach- und Informationswissenschaften - warum und wie?

Handreichung:

Empfehlungen zu datentechnischen Standards und Tools bei der Erhebung von Sprachkorpora

Inhalt

Vorbemerkungen	2
Einleitung	2
Teil 1: Empfehlungen zu datentechnischen Standards und Tools bei der Erhebung mündlicher Korpora	4
1.1. Primärdaten	4
1.1.1. Audio- und Videoaufnahmen	4
1.1.2. Zusatzmaterialien	6
1.2. Transkription und weitere Annotation	7
1.2.1. Tools und Formate	7
1.2.2. Transkriptionskonventionen und Annotationsschemata	8
1.3. Metadaten	10
1.4. Archivierung	12
Referenzen zu Teil 1	14
Teil 2: Empfehlungen zu datentechnischen Standards und Tools bei der Erhebung von Schriftkorpora	17
2.1 Digitalisierung und Korpuserfassung	17

Universitätsbibliothek Hildesheim:

Workshop: Forschungsdatenmanagement in den Sprach- und Informationswissenschaften - warum und wie?

Zusammenfassung:

- es gibt ein breites Spektrum an FD
- es gibt weiterhin viele Herausforderungen und Desiderate
- die FD in den Sprach- und Informationswissenschaften sind von dauerhafter Relevanz für die Forschung

Love your data!