

Data for my research: Where can I get it, where do I take it, what can I do with it, how do I use it in my resume?

Looking at the Research Data Infrastructures CLARIN in Europe and Text+ in Germany

Thorsten Trippel, Leibniz-Institut für Deutsche Sprache,
Mannheim
trippel@ids-mannheim.de

- Development and work by the Text+ community
- ... And the CLARIN community



THORSTEN TRIPPEL Computational Linguist | Ph.D.

CONTACT INFORMATION: Email: thorsten.trippel@email.com Phone: +49 1234567890 Address: 123 Research Street, Tübingen, Germany

EDUCATION: Ph.D. in Computational Linguistics, Bielefeld University, Bielefeld, Germany, 20XX-20XX

- Dissertation: "Exploring Language Processing Models for Sentiment Analysis"
- Advisor: Prof. Anna Müller

M.A. in Computational Linguistics, Bielefeld University, Bielefeld, Germany, 20XX-20XX

- Thesis: "Statistical Approaches to Text Classification"
- Advisor: Prof. Hans Schmidt

B.A. in Linguistics, Bielefeld University, Bielefeld, Germany, 20XX-20XX

RESEARCH INTERESTS:

- Research Data Management
- Graph Representations of Lexical Resources
- Standardization in Linguistics

EXPERIENCE: Researcher, University of Tübingen, Tübingen, Germany, 20XX-present

- Leading research projects on research data management graph representations of lexical resources, and standardization in linguistics
- Designing and implementing frameworks for managing linguistic datasets and ensuring data quality
- Collaborating with international partners to develop standards for linguistic resources

Project Leader, ISO 24622-2: Lexical Markup Framework (LMF), International Organization for Standardization (ISO), 20XX-present

- Leading the development and standardization of ISO 24622-2, focusing on graph-based representations of lexical resources
- Coordinating international working groups and stakeholders to define data models and exchange formats
- Promoting the adoption and implementation of the standard within the linguistic community

Research Associate, Leibniz-Institut of German Language (IDS), Mannheim, Germany, 20XX-present

- Conducting research on graph-based representations of lexical resources and their applications in computational linguistics
- Analyzing and processing large-scale linguistic datasets using graph databases
- Publishing research findings in academic journals and presenting at conferences

Teaching Assistant, Bielefeld University, Bielefeld, Germany, 20XX-20XX

- Assisted in teaching undergraduate courses on computational linguistics and natural language processing
- Mentored and supervised student research projects
- Graded assignments and provided constructive feedback to students

PUBLICATIONS:

- Trippel, T., Müller, A. (20XX). "A Comparative Study of Sentiment Analysis Techniques on Social Media Data." *Journal of Computational Linguistics*, 25(2), 123-145.
- Trippel, T., Schmidt, H. (20XX). "Statistical Approaches to Text Classification: A Review." *Proceedings of the International Conference on Natural Language Processing*, 345-356.
- Trippel, T., Müller, A., Schmidt, H. (20XX). "Graph-Based Representation of Lexical Resources: An Overview." *Journal of Language Resources and Standardization*, 30(4), 567-589.

THORSTEN TRIPPEL Computational Linguist | Ph.D.

CONTACT INFORMATION: Email: thorsten.trippel@email.com Phone: +49 1234567890 Address: 123 Research Street, Tübingen, Germany

EDUCATION: Ph.D. in Computational Linguistics, Bielefeld University, Bielefeld, Germany, 20XX-20XX

- Dissertation: "Exploring Language Processing Models for Sentiment Analysis"
- Advisor: Prof. Anna Müller

M.A. in Computational Linguistics, Bielefeld University, Bielefeld, Germany, 20XX-20XX

- Thesis: "Statistical Approaches to Text Classification"
- Advisor: Prof. Hans Schmidt

B.A. in Linguistics, Bielefeld University, Bielefeld, Germany, 20XX-20XX

RESEARCH INTERESTS:

- Research Data Management
- Graph Representations of Lexical Resources
- Standardization in Linguistics

EXPERIENCE: Researcher, University of Tübingen

- Leading research projects on research data management, graph-based representations of lexical resources, and standardization
- Designing and implementing frameworks for data management, ensuring data quality
- Collaborating with international partners to develop and share lexical resources

Project Leader, ISO 24622-2: Lexical Markup Framework (LMF) Organization for Standardization (ISO), 20XX-present

- Leading the development and standardization of ISO 24622-2: graph-based representations of lexical resources
- Coordinating international working groups and stakeholders to define data models and exchange formats
- Promoting the adoption and implementation of the standard within the linguistic community

Research Associate, Leibniz-Institut of German Language (IDS), Mannheim, Germany, 20XX-present

- Conducting research on graph-based representations of lexical resources and their applications in computational linguistics
- Analyzing and processing large-scale linguistic datasets using graph databases
- Publishing research findings in academic journals and presenting at conferences

Teaching Assistant, Bielefeld University, Bielefeld, Germany, 20XX-20XX

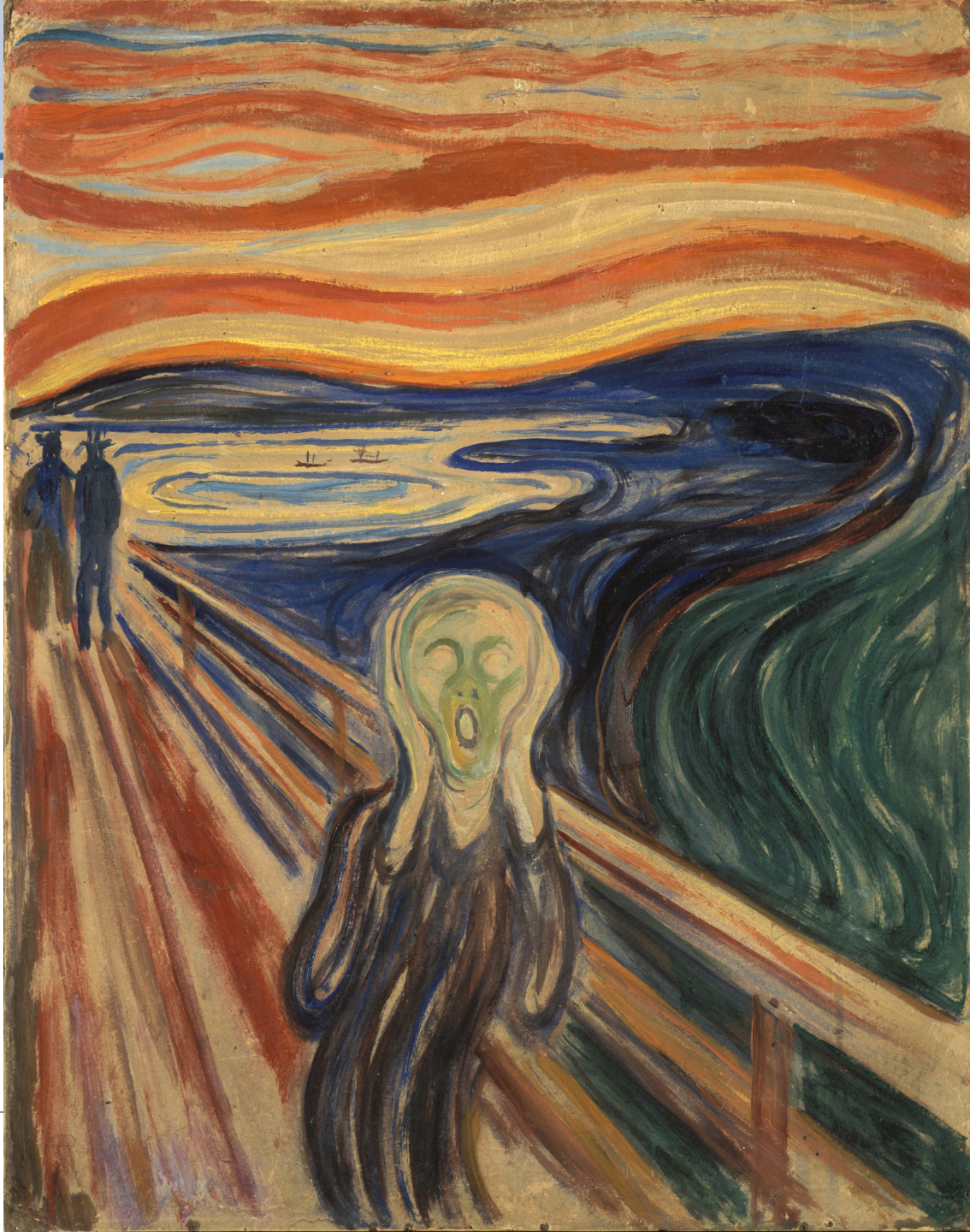
- Assisting in teaching undergraduate courses on computational linguistics and natural language processing
- Supervising and mentoring student research projects
- Providing constructive feedback to students

Journal Article: "A Comparative Study of Sentiment Analysis" Journal of Computational Linguistics, 25(2), 20XX

Conference Paper: "Statistical Approaches to Text Classification: A Case Study" International Conference on Natural Language Processing, 20XX

Book Chapter: "Graph-Based Representation of Lexical Resources and their Applications" Computational Linguistics and Natural Language Processing, 20XX

Something between "Not really" and "really not"



[The Scream](#)
[\(MM.M.00514\)](#) by
[Edvard Munch, 1910](#)
[https://commons.wikimedia.org/wiki/File:
Edvard_Munch_-
_The_Scream_-
_Google_Art_Project.
jpg](https://commons.wikimedia.org/wiki/File:Edvard_Munch_-_The_Scream_-_Google_Art_Project.jpg)

- Reproducibility crisis
- Empirical turn

- Reproducibility crisis
- Empirical turn
- ChatGPT



Can you make bullet points out of that?



Sure, here are the main points summarized as bullet points:

- The reproducibility crisis refers to the problem of many scientific studies being difficult or impossible to reproduce.
- The crisis is fueled by factors such as the pressure to publish, selective reporting of results, underpowered studies, and cultural issues in the scientific community.
- To address the reproducibility crisis, proposed solutions include greater transparency, open data, pre-registration of studies, and improved statistical training for researchers.
- Addressing this multifaceted issue will require a concerted effort from scientists, publishers, funders, and institutions to promote a more rigorous collaborative scientific culture.

- The reproducibility crisis refers to the problem of many scientific studies being **difficult or impossible to reproduce.**
- The crisis is fueled by factors such as the pressure to publish, selective reporting of results, underpowered studies, and cultural issues in the scientific community.
- To address the reproducibility crisis, proposed solutions include **greater transparency, open data**, pre-registration of studies, and **improved statistical training for researchers.**
- Addressing this multifaceted issue will require a concerted effort from scientists, publishers, funders, and institutions to promote a more rigorous and collaborative scientific culture.



Sure, here are the main points of the empirical turn in the humanities and social sciences summarized as a bulleted list:

- The empirical turn is a growing trend in the humanities and social sciences towards using empirical research methods to study human behavior and social phenomena.
- Factors driving the empirical turn include the increasing availability of data, especially with the rise of digital technology and the internet.
- Empirical research methods can provide more objective and quantitative measures of human behavior and social phenomena, and can help to identify patterns and relationships that may not be immediately apparent through more qualitative methods.
- However, there are also criticisms of the empirical turn, particularly with regards to the potential loss of focus on more interpretive and subjective approaches to studying human behavior and social phenomena.
- Additionally, there are concerns about the potential misuse or misinterpretation of data, particularly in fields where there is a lack of standardized methods for collecting and analyzing data.
- The empirical turn represents an ongoing debate about the role of empirical research methods in the humanities and

Example for empirical turn and reproducibility crisis: ChatGPT

- » ChatGPT:
 - » large language model
 - » Training involves large sets of language data
 - » Language data is research data in the humanities – and in AI
 - » Empirical Research (!)
- » Asking twice not always provides the same answers
- » Where is the source of ChatGPT's "facts"? Who gets credited for the actual research?

of data, particularly in fields where there is a lack of standardized methods for collecting and analyzing data.

- The empirical turn represents an ongoing debate about the role of empirical research methods in the humanities and

Example for empirical turn reproducibility crisis: ChatGPT

- » ChatGPT:
 - » large language model
 - » Training involves
 - » Language models generate text – and in AI
 - »
- » ChatGPT generates the same answers
- » ChatGPT's "facts"? Who gets credited for the actual

Honesty?
AI driven plagiarism?

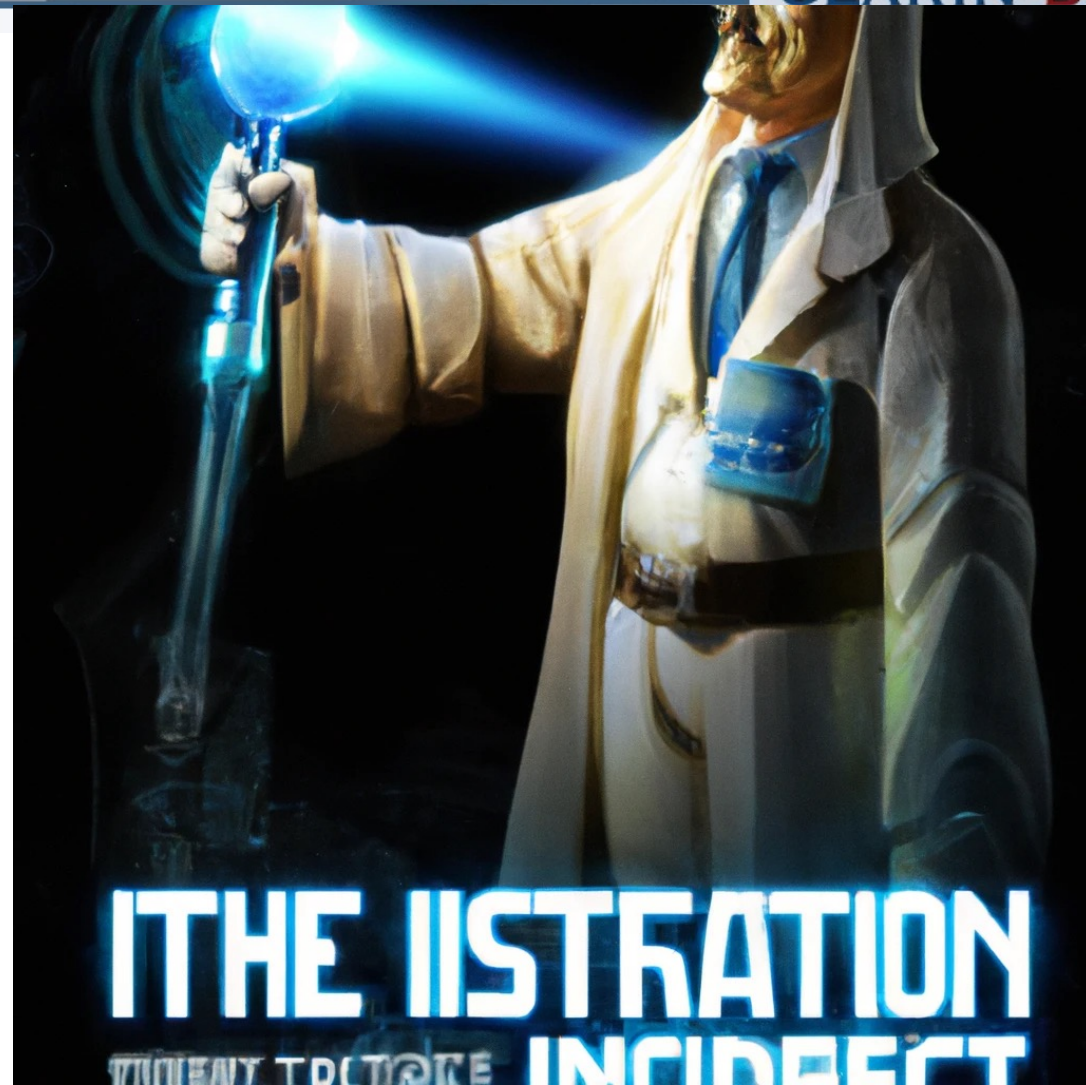
of data, particularly in fields where there is a lack of standardized methods for collecting and analyzing data.

- The empirical turn represents an ongoing debate about the role of empirical research methods in the humanities and

cial
s
social
a,
tify
more
station

The return of integrity

- Coming soon?
- Produced by DALL-E, with the prompt
Generate a film poster for "Return of Integrity"
in the style of a Star Wars poster. include a
scholar, a microscope, a manuscript and a
microphone
<https://labs.openai.com/>



- Archiving empirical data
- Reproducing statistical results by storing the formulae and data
- Making the data accessible

- Research questions based on empirical information require data
- Digital Methods allow processing data that cannot be processed without them (in a reasonable time)
- Data needs to be found, collected, evaluated and at least pre-processed to be available for further processing

- Data mining and data analysis require digital/digitized data
 - Depending on the methods applied, the data needs structures
 - There is no such thing as unstructured data – sometimes the structures may be hidden
 - Data analysis depends on the research question

- Providing research results based on empirical data requires some form of visualization.
- Figures and tables are typical forms of visualizing data
- Textual descriptions may also be a form of visualization

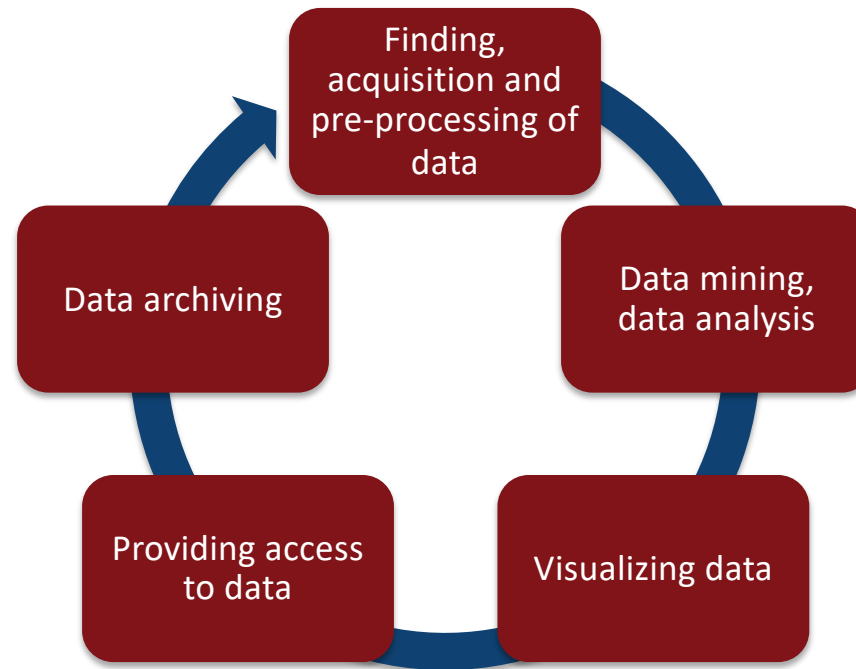
- Providing research results based on empirical data requires some form of visualization.
- Figures and tables are typical forms of visualizing data
- Textual descriptions may also be a form of visualization

mmh

- Reproducibility of results is a quality indicator
 - Reviewers need access to underlying data
- Data may be reused for different research purposes
 - Dual use (!)
 - Alternative use
 - Later use
 - Meta use

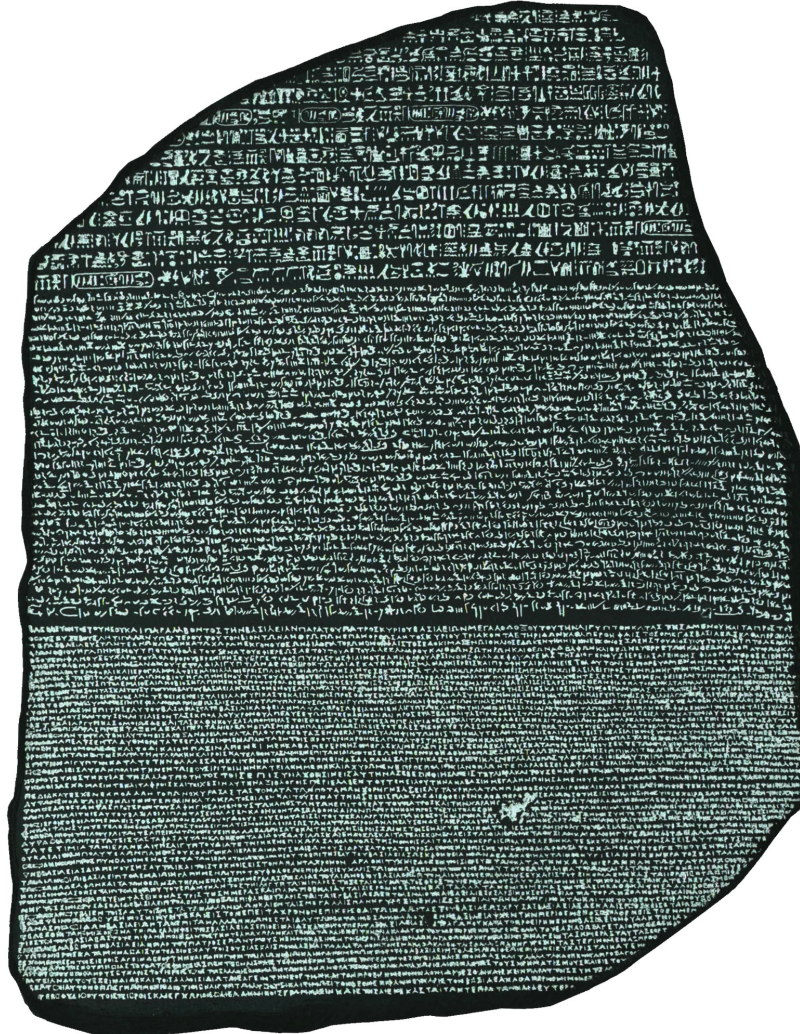
- Hypothesis: same source “situation” \Rightarrow same results
 - BUT: statistical deviations
 - Non-reproducible events
 - (Technical) deviations in the analysis, e.g. interpretation
- Hypothesis: Reproducibility is a sign of quality
 - (see above)

- Long term access to data requires archiving providing defined access to the data
- Scholars will not be able to sustainably provide access to the data themselves.





https://upload.wikimedia.org/wikipedia/commons/c/ca/Rosetta_Stone_BW.jpeg



- Findable
 - Accessible
 - Interoperable
 - Reusable
- » Collective Benefit
 - » Authority to control
 - » Responsibility
 - » Ethics

Wilkinson, Mark D et al. (2016). "The FAIR Guiding Principles for scientific data management and stewardship". In: Scientific data 3.

Carroll, Stephanie Russo, Edit Herczog, Maui Hudson, Keith Russell, and Shelley Stall (Apr. 2021). "Operationalizing the CARE and FAIR Principles for Indigenous data futures". In: Scientific Data 8.1, p. 108. url: <https://doi.org/10.1038/s41597-021-00892-0> .

1. To be **Findable** any Data Object should be uniquely and persistently identifiable
 - The same Data Object should be re-findable at any point in time, thus Data Objects should be **persistent**, with emphasis on their metadata,
 - A Data Object should minimally contain basic machine actionable metadata that allows it to be distinguished from other Data Objects
 - Identifiers for any concept used in Data Objects should therefore be **Unique** and **Persistent**

See: <https://www.force11.org/node/6062>

2. Data is **Accessible** in that it can be always obtained by machines and humans
- Upon appropriate authorization
 - Through a well-defined protocol
 - Thus, machines and humans alike will be able to judge the actual accessibility of each Data Object.

See: <https://www.force11.org/node/6062>

- ### 3. Data Objects can be **Interoperable** only if:
- (Meta) data is machine-actionable
 - (Meta) data formats utilize shared vocabularies and/or ontologies
 - (Meta) data within the Data Object should thus be both syntactically parseable and semantically machine-accessible

See: <https://www.force11.org/node/6062>

4. For Data Objects to be **Re-usable** additional criteria are:

- Data Objects should be compliant with **principles 1-3**
- (Meta) data should be sufficiently well-described and rich that it can be automatically (or with minimal human effort) linked or integrated, like-with-like, with other data sources
- Published Data Objects should refer to their sources with rich enough metadata and provenance to enable proper citation.

See: <https://www.force11.org/node/6062>

- Non reproducible data
 - Amount of data/work involved for creation or collection of data
 - Non reproducible events
- Reproducible data
 - "recipe" including technical environments, selection criteria, (acceptable) statistical variance, ...

The CARE Principles

○ Indigenous Data Governance Principles



Collective Benefit

Data ecosystems shall be designed and function in ways that enable Indigenous Peoples to derive benefit from the data.

C1

For inclusive development and innovation

Governments and institutions must actively support the use and reuse of data by Indigenous nations and communities by facilitating the establishment of the foundations for Indigenous innovation, value generation, and the promotion of local self-determined development processes.

C2

For improved governance and citizen engagement

Data enrich the planning, implementation, and evaluation processes that support the service and policy needs of Indigenous communities. Data also enable better engagement between citizens, institutions, and governments to improve decision-making. Ethical use of open data has the capacity to improve transparency and decision-making by providing Indigenous nations and communities with a better understanding of their peoples, territories, and resources. It similarly can provide greater insight into third-party policies and programs affecting Indigenous Peoples.

C3

For equitable outcomes

Indigenous data are grounded in community values, which extend to society at large. Any value created from Indigenous data should benefit Indigenous communities in an equitable manner and contribute to Indigenous aspirations for wellbeing.

https://www.gida-global.org/s/CARE-Principles_One-Pagers-FINAL_Oct_17_2019.pdf

•

- Pay and credit
- Publish results (Zero results? What if the publication has a negative impact?)
- Privacy law (GDPR, but publication?)
- Respect ethics and legal obligations

Tell me more: How?



- Common Language Resources and Technology Infrastructure
- ESFRI Project
- Target group: Humanities scholars and social scientists working in language-based research
 - Language as a research tool and as a research object

- Austria
 - Belgium
 - Bulgaria
 - Croatia
 - Cyprus
 - Czech Republic
 - Denmark
 - Estonia
 - Finland
- » Germany
- » Greece
- » Hungary
- » Iceland
- » Italy
- » Latvia
- » Lithuania
- » The Netherlands
- » Norway
- » Poland
- » Portugal
- » Slovenia
- » Sweden
- 
- South Africa
 - Switzerland
 - United Kingdom
 - Carnegie Mellon University (USA)

The situation in Germany



Based on a decision by the federal government and the states in 2018: Construction of a German national research data Infrastructure

Aim: ensuring access and long time archiving of data resources

- » Implementing the FAIR principles
- » Discipline independent and not bound to individual states

“The NFDI is being established in a science-driven process as a networked structure of consortia acting on their own initiative.”

“Die NFDI wird in einem aus der Wissenschaft getriebenen Prozess als vernetzte Struktur eigeninitiativ agierender Konsortien aufgebaut.”

https://www.dfg.de/foerderung/info_wissenschaft/2020/info_wissenschaft_20_29/index.html

Credits: <https://www.pexels.com/de-de/foto/dock-unter-der-goldenen-stunde-276259/>



Data as a common asset for excellent research, organized by research in Germany.

GEFÖRDERT VOM



Bundesministerium
für Bildung
und Forschung



Step by step, we are improving the possibilities for using data for science and society. Through our collaboration in the NFDI Association, an umbrella **organization for research data management** in all branches of science is emerging. In **collaboration with national and international partners**, we are creating the framework for **legally compliant, interoperable and sustainable data infrastructures** that are readily accessible to researchers in their everyday work. We train, strengthen competence in handling data, and open up new career paths.

Text- and language based research data are at the centre of the Text+ research data infrastructure within initiative to establish a National Research Data Infrastructure (NFDI). Text+ focusses on three data domains: collections, lexical resources, and editions.

Focus:

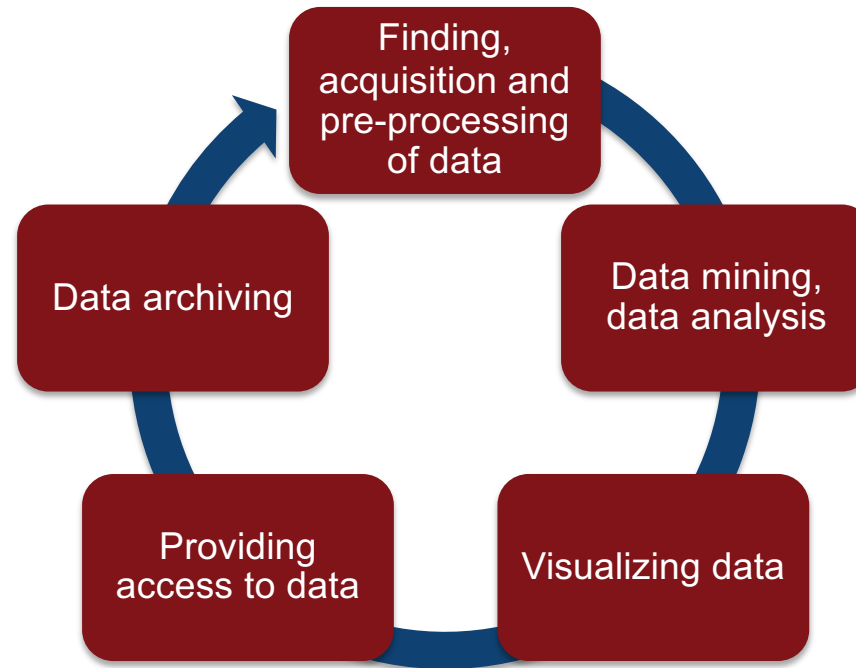
COLLECTIONS

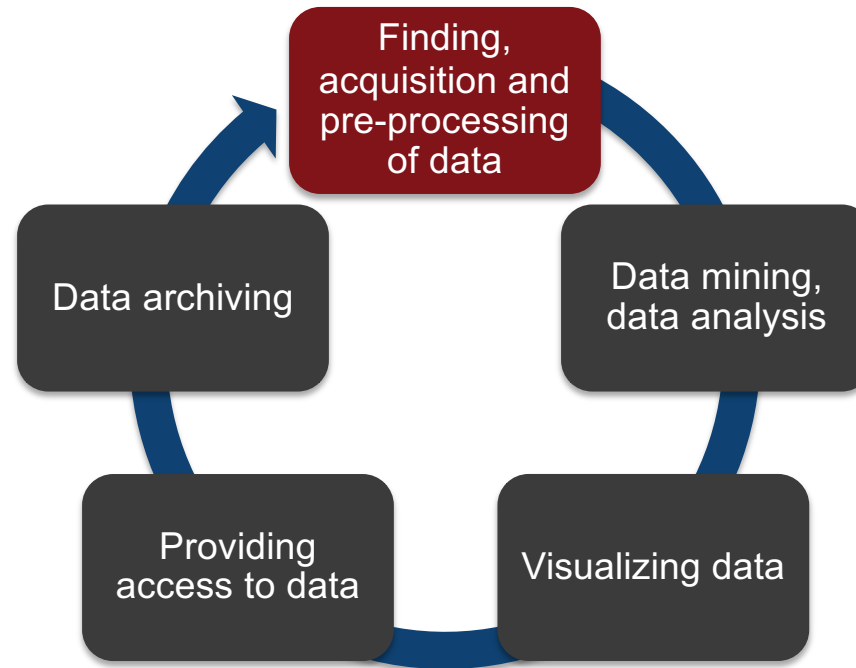
LEXICAL
RESOURCES

Editions

- Located at different institutions
- Legacy data sets
- Data formats
- Query interfaces

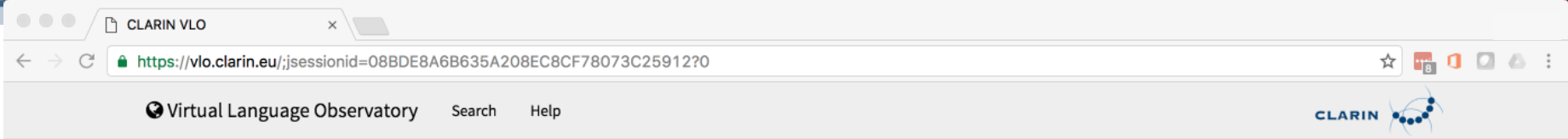








VLO: Virtual Language Observatory – A search engine for Research data supporting Findability



CLARIN Virtual Language Observatory

Use the search bar to start searching the VLO, or continue to browse everything and use facets to narrow down to your area of interest or discover new resources.

[See all records](#)

Search



Showing all 909552 records

Results per page: 10

Use the categories below to limit the search results to those matching the selected value(s).

Language

Collection

Resource type

Modality

Format

<< < 1 2 3 4 5 6 7 8 9 10 > >>

ukr_newscrawl_2011_1M

1.000.000 sentences of a Ukrainian newspaper corpus based on material crawled in 2011



rus_news_2013_1M

1.000.000 sentences of a Russian newspaper corpus based on material from 2013



deu_news_2012_3M

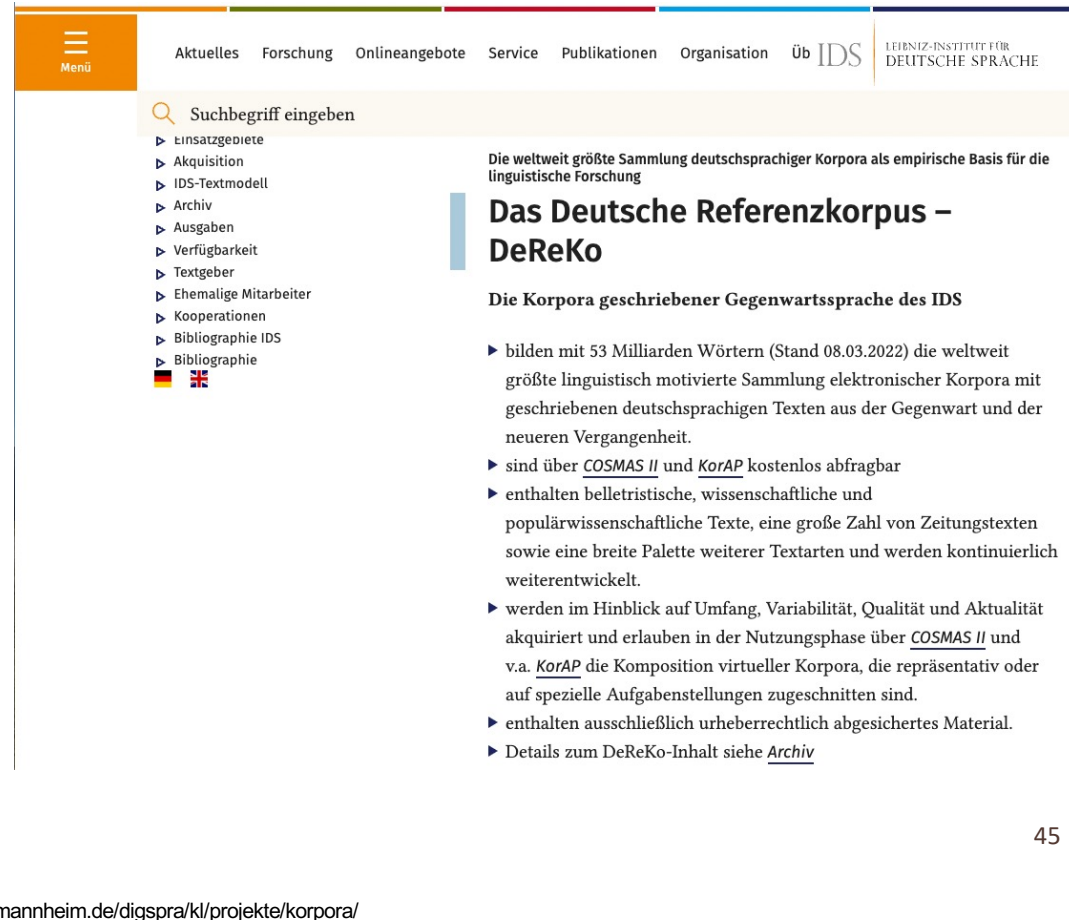
3.000.000 sentences of a German newspaper corpus based on material from 2012



- » Language Resources
- » Modern German
- » 53 Billion words
- » Collected and maintained since the 1960's
- » Newspapers, political speeches, books, fairy tales,
- » Includes copyright protected resources
- » Host: IDS Mannheim

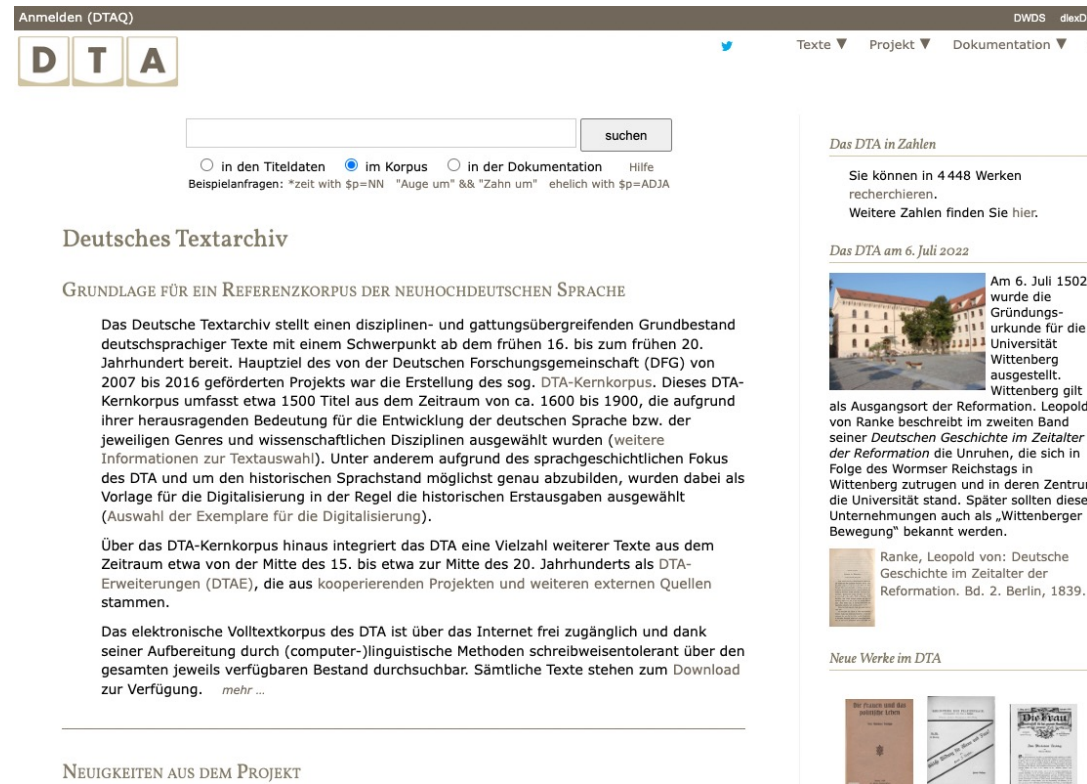


<https://www.ids-mannheim.de/digspra/kl/projekte/korpora/>



The screenshot shows the website for the DeReKo project. The header includes a navigation menu with items like 'Aktuelles', 'Forschung', 'Onlineangebote', 'Service', 'Publikationen', 'Organisation', and 'IDS'. Below the header is a search bar with the text 'Suchbegriff eingeben'. A dropdown menu is visible, listing various categories such as 'Einsatzgebiete', 'Akquisition', 'IDS-Textmodell', 'Archiv', 'Ausgaben', 'Verfügbarkeit', 'Textgeber', 'Ehemalige Mitarbeiter', 'Kooperationen', 'Bibliographie IDS', and 'Bibliographie'. The main content area features a section titled 'Das Deutsche Referenzkorpus – DeReKo' with a sub-heading 'Die Korpora geschriebener Gegenwartssprache des IDS'. Below this, there is a list of bullet points describing the corpus: it is the largest collection of German corpora, contains over 53 billion words, is available for free via COSMAS II and KORAP, and includes a wide range of text types from newspapers to scientific texts. A QR code is located at the bottom left of the screenshot, and the URL <https://www.ids-mannheim.de/digspra/kl/projekte/korpora/> is displayed at the bottom.

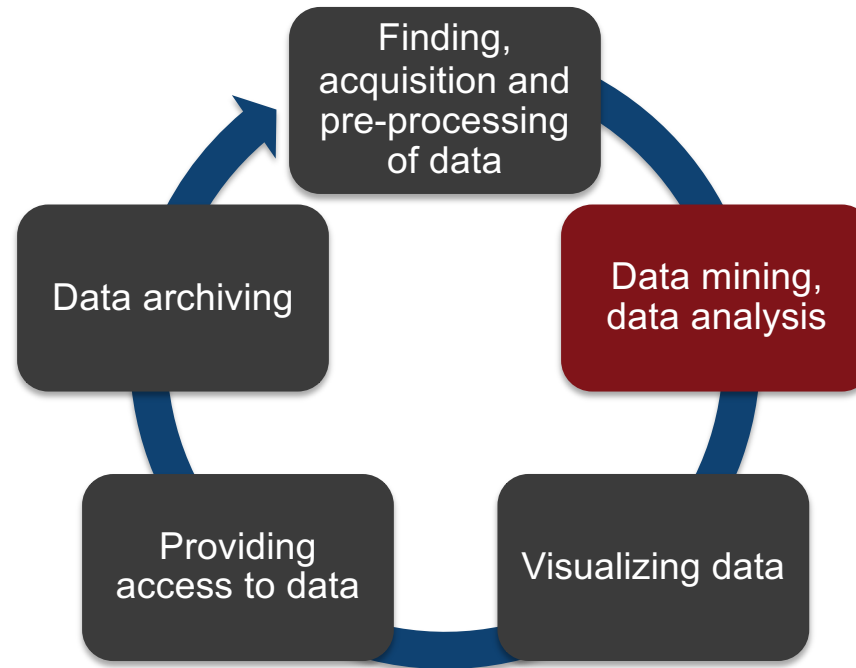
- » German texts from the 16th to the 20th century
- » 1500 titles
- » Available in different formats, including raw text, TEI XML, ...
- » Mostly open CC-BY-SA licenced
- » Host: BBAW Berlin



The screenshot shows the homepage of the German Text Archive (DTA). At the top, there is a navigation bar with 'Anmelden (DTAQ)', 'DWDS', and 'dlexDB'. Below this is a search bar with a 'suchen' button and radio buttons for 'in den Titeldaten', 'im Korpus' (selected), and 'in der Dokumentation'. A search example is provided: '*zeit with \$p=NN "Auge um" && "Zahn um" ehelich with \$p=ADJA'. The main heading is 'Deutsches Textarchiv' with the subtitle 'GRUNDLAGE FÜR EIN REFERENZKORPUS DER NEUHOCHDEUTSCHEN SPRACHE'. The text describes the archive's mission and the DTA-Kernkorpus. A QR code is visible in the bottom right of the screenshot area.

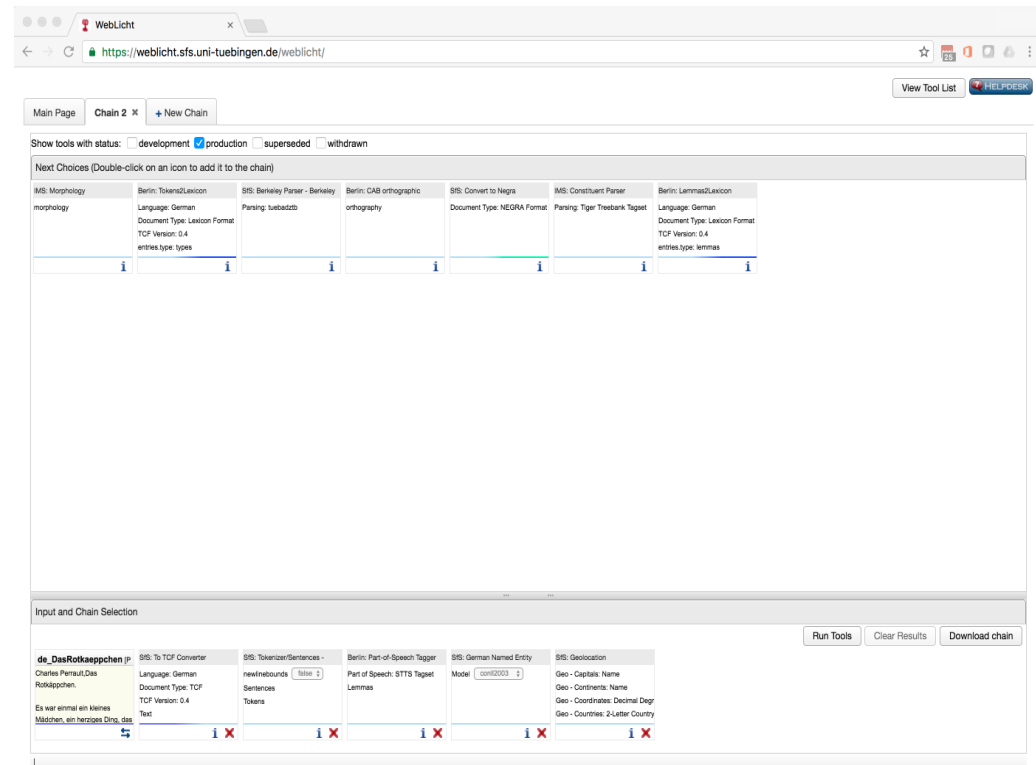


<https://www.deutschestextarchiv.de/>

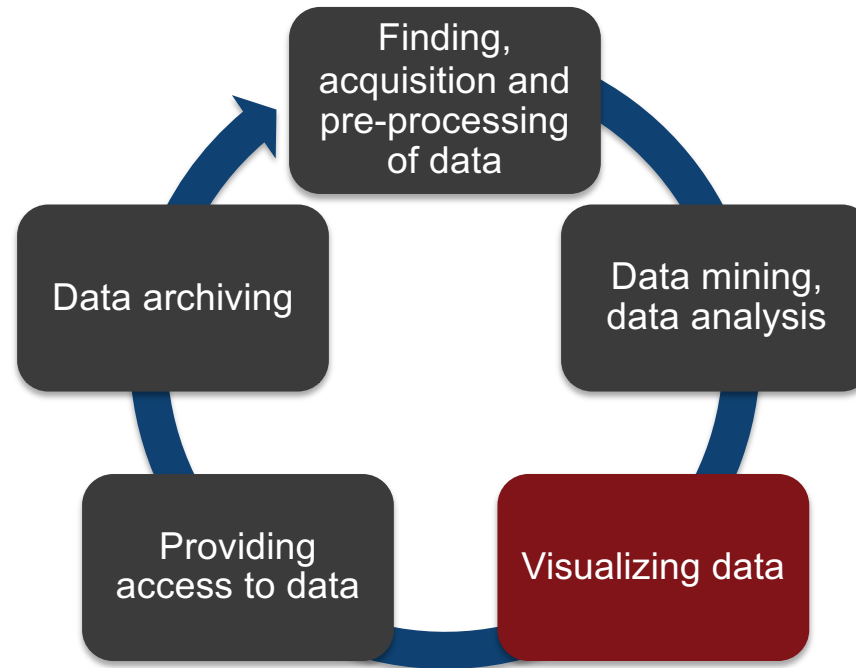


Virtual research environment for text annotation: WebLicht

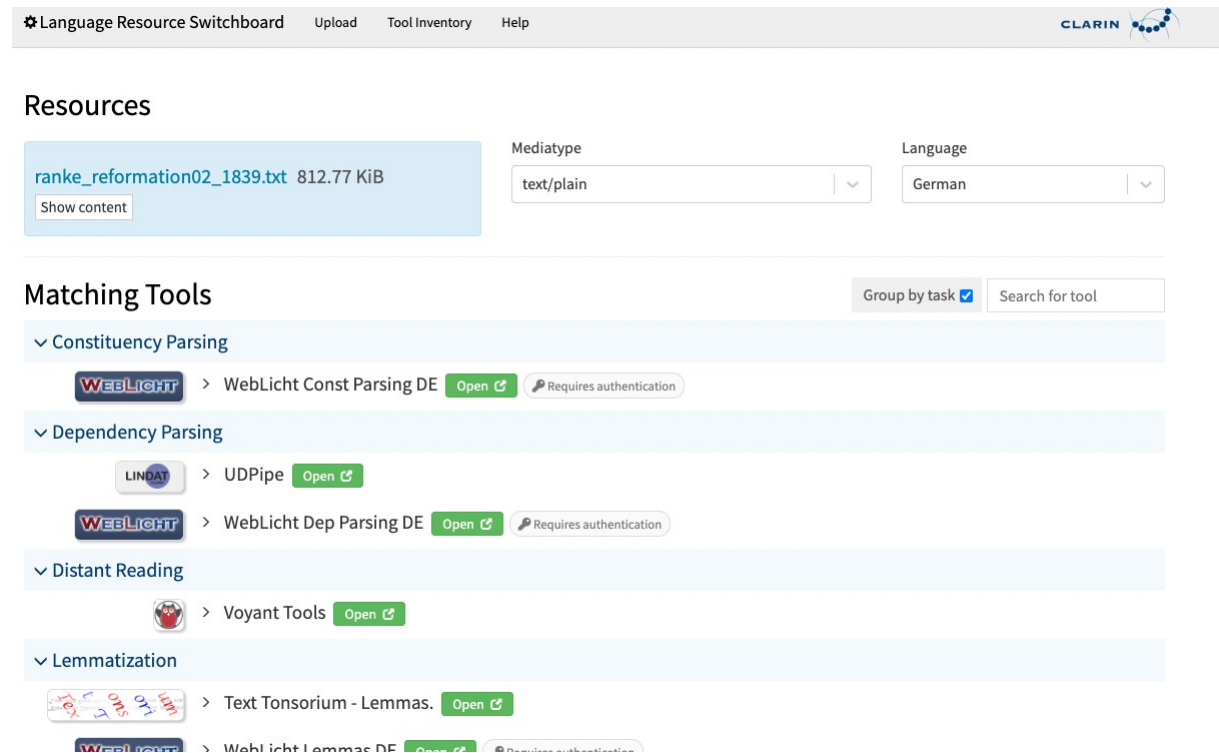
- » Webapplication with various essential tools when working with language data, including
 - » Tokenizer
 - » Parser
 - » Named Entity Recognizer
- » Multi-language support
- » Academic login required
- » Host: University of Tübingen



<https://weblicht.sfs.uni-tuebingen.de/>



- » For example integrated in DTA
- » Here: via reference to the Language Resource Switchboard
- » Reference to many other tools

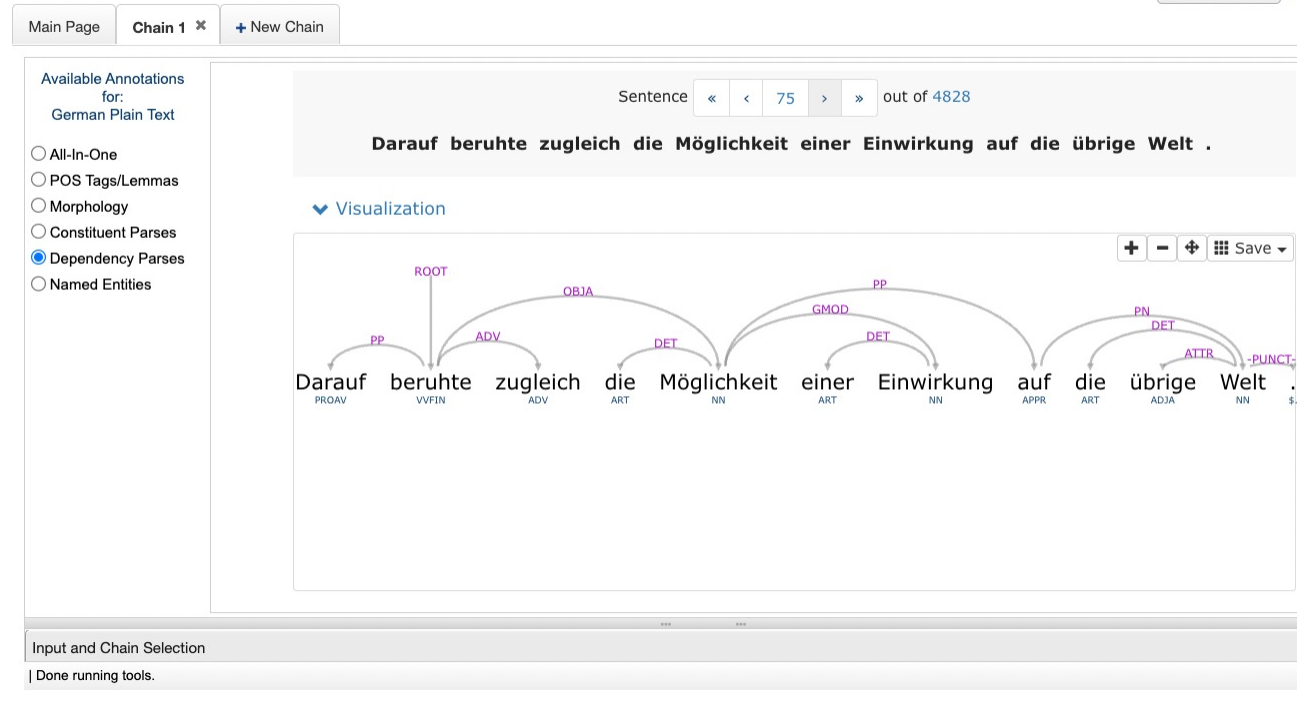


The screenshot shows the Language Resource Switchboard interface. At the top, there are navigation links: "Language Resource Switchboard", "Upload", "Tool Inventory", and "Help". The "Resources" section displays a file named "ranke_reformation02_1839.txt" (812.77 KiB) with a "Show content" button. To the right, there are dropdown menus for "Mediatype" (set to "text/plain") and "Language" (set to "German"). The "Matching Tools" section is organized into categories: "Constituency Parsing" (with "WebLicht Const Parsing DE" and "Requires authentication"), "Dependency Parsing" (with "UDPipe" and "WebLicht Dep Parsing DE" and "Requires authentication"), "Distant Reading" (with "Voyant Tools"), and "Lemmatization" (with "Text Tonsorium - Lemmas" and "WebLicht Lemmas DE" and "Requires authentication"). A search bar and a "Group by task" checkbox are also visible.

https://switchboard.clarin.eu/#/vlo/http%3A%2F%2Fwww.deutschestextarchiv.de%2Fbook%2Fdownload_txt%2Ffranke_reformation02_1839%2Ftext%2Fplain



- » Integrated in Virtual Research Environment WebLicht
- » For searching and visualizing syntactic annotations
- » Host: University of Tübingen

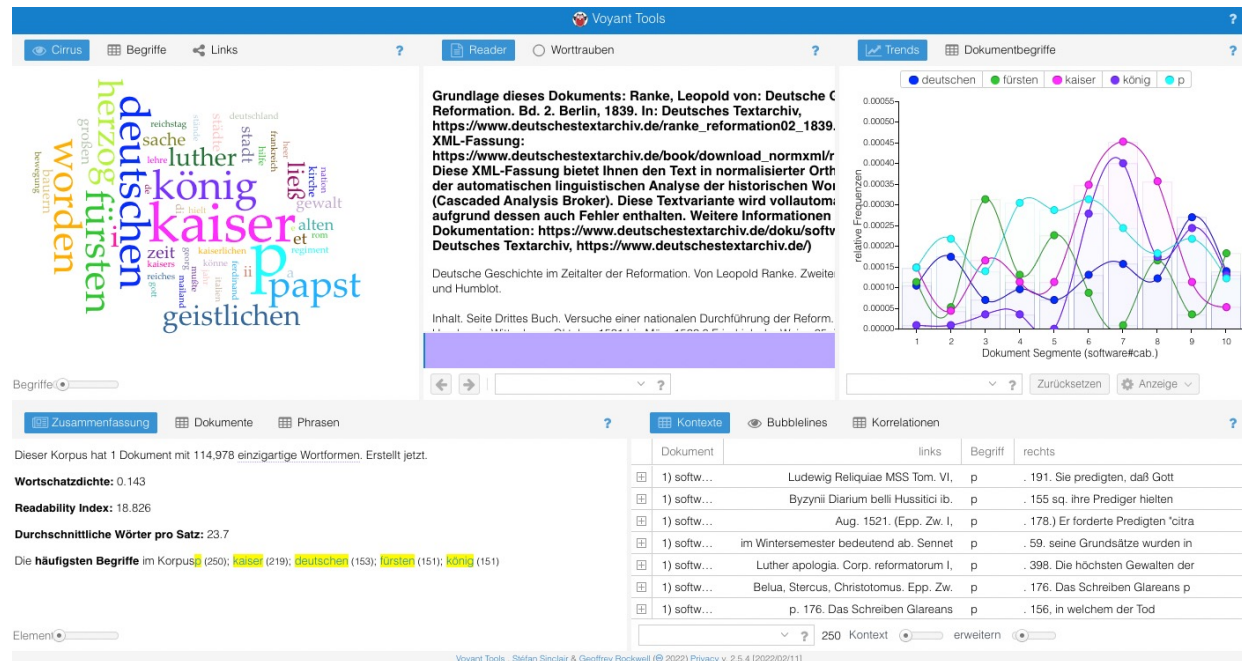


The screenshot shows the Tüandra web interface. At the top, there are navigation tabs: 'Main Page', 'Chain 1 ✕', and '+ New Chain'. Below this, a sidebar lists 'Available Annotations for: German Plain Text' with radio buttons for 'All-In-One', 'POS Tags/Lemmas', 'Morphology', 'Constituent Parses', 'Dependency Parses' (selected), and 'Named Entities'. The main area displays the sentence 'Darauf beruhte zugleich die Möglichkeit einer Einwirkung auf die übrige Welt .' with a dependency parse graph. The graph shows nodes for each word with their POS tags and lemmas below them, and arcs connecting them with dependency labels like 'ROOT', 'PP', 'OBJA', 'ADV', 'DET', 'GMOD', 'PP', 'PN', 'DET', 'ATTR', and 'PUNCT'. A 'Visualization' section is expanded. At the bottom, there is an 'Input and Chain Selection' section with the text '| Done running tools.'.



<https://weblicht.sfs.uni-tuebingen.de/weblicht/>

- » For example integrated in DTA
- » Here: Voyant-tools

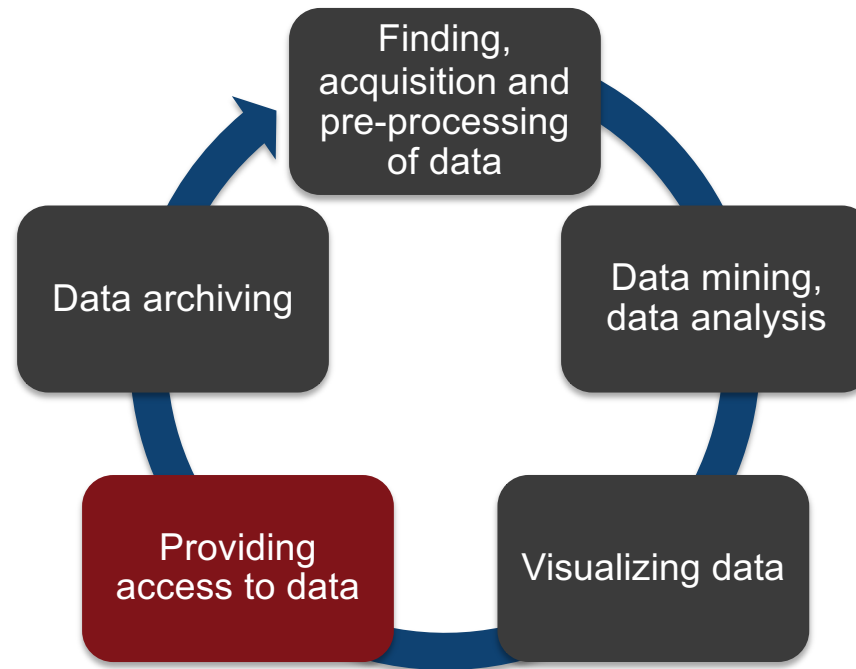


The screenshot displays the Voyant Tools interface. On the left, a word cloud features prominent terms like 'deutsch', 'fürsten', 'kaiser', 'papst', and 'geistlichen'. The central text area shows a snippet from 'Grundlage dieses Dokuments: Ranke, Leopold von: Deutsche Reformation, Bd. 2. Berlin, 1839...' with a highlighted section. On the right, a line graph plots 'relative Frequenzen' across 'Dokument Segmente (software#cab.)' for categories 'deutschen', 'fürsten', 'kaiser', 'könig', and 'p'. The bottom section shows a 'Kontexte' table with columns for 'Dokument', 'links', 'Begriff', and 'rechts'.

Dokument	links	Begriff	rechts
1) softw...	Ludewig Reliquiae MSS Tom. VI.	p	. 191. Sie predigten, daß Gott
1) softw...	Byzynth Diarium belli Hussitici ib.	p	. 155 sq. ihre Prediger hielten
1) softw...	Aug. 1521. (Epp. Zw. I.	p	. 178.) Er forderte Predigten 'citra
1) softw...	im Wintersemester bedeutend ab. Sennet	p	. 59. seine Grundsätze wurden in
1) softw...	Luther apologia. Corp. reformatum I.	p	. 398. Die höchsten Gewalten der
1) softw...	Belua, Stercus, Christotomus. Epp. Zw.	p	. 176. Das Schreiben Glareans p
1) softw...	p. 176. Das Schreiben Glareans	p	. 156. in welchem der Tod



<https://voyant-tools.org>



- » For reference in publications and citation of data
- » Allows readers to receive (information on) access to the data
- » Example:
 - » Handle:
<https://hdl.handle.net/11022/0000-0007-CFE2-1>
 - » Persistent URLs:
https://www.deutschestextarchiv.de/ranke_reformamtion02_1839
 - » DOIs: [10.5281/zenodo.3735141](https://doi.org/10.5281/zenodo.3735141)

TALAR - Tübingen Archive of Language Resources

About Contact Data Management Help Privacy Statistics Technology

Resource: Data associated with Dima (2019) dissertation: "Composition Models for the Representation and Semantic Interpretation of Nominal Compounds"

General Info Project Publications Creation Access **Resource-specific information** Data files About...

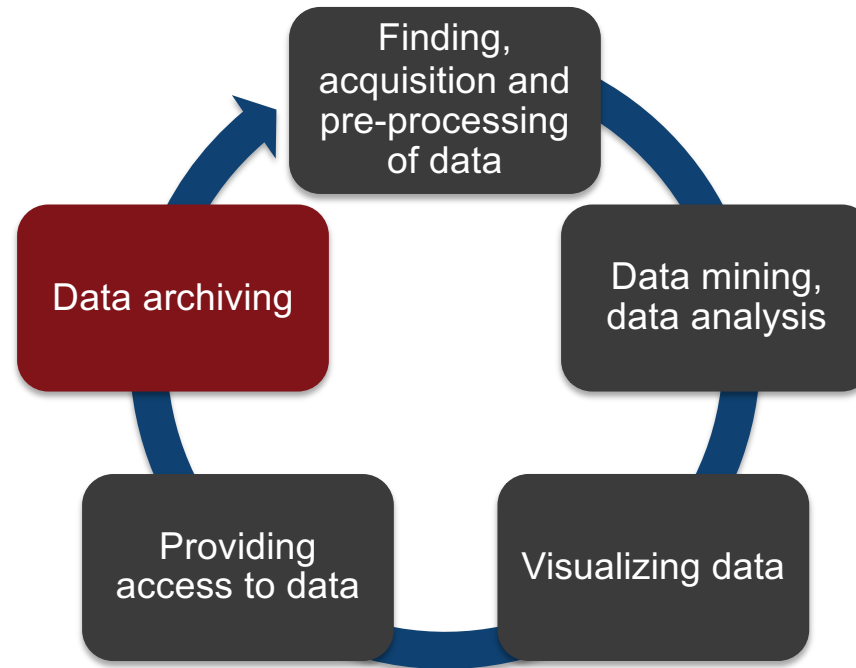
Cite as

Dima C. (2019): Data associated with Dima (2019) dissertation: "Composition Models for the Representation and Semantic Interpretation of Nominal Compounds" Persistent identifier: <https://hdl.handle.net/11022/0000-0007-CFE2-1>

This resource is provided through the technology partnership with the Tübingen Archive of Language Resources

URL zu diesem Werk: https://www.deutschestextarchiv.de/ranke_reformation02_1839
URL zu dieser Seite: https://www.deutschestextarchiv.de/ranke_reformation02_1839/18
Zitationshilfe: Ranke, Leopold von: Deutsche Geschichte im Zeitalter der Reformation
<https://www.deutschestextarchiv.de/ranke_reformation02_1839/18>

Blaette, Andreas (2020): GermaParl. Linguistically Annotated and Indexed Corpus of Plenary Protocols of the German Bundestag. CWB corpus version 1.0.6. <https://doi.org/10.5281/zenodo.3735141>



- » Distributed over many institutions
 - » According to their specialization
 - » Processes to allow depositing of data of third parties
 - » All partners mentioned hosting data for the community
- » Open to new partners
 - » Reliability (certification)
 - » Relevance to the Text+ domains
 - » Standard interfaces
- » Fall back: bitstream preservation

- If you produce data
 - Store it in a credible repository with a persistent identifier
- If you used data
 - Cite it with the persistent identifier
- If you analysed data
 - Refer to the tools with a persistent identifier
- List your data on your CV just as any other citation!

- If you produce data
 - Store it in a credible repository with a persistent identifier
- If you used data
 - Cite it with the persistent identifier
- If you analyse data
 - Refer to the tool with a persistent identifier
- List your data on your CV just as any other citation!

Follow good scientific practices!

How can I get my data on my CV?

- If you produce data
 - Store it in a **CR** with a persistent identifier
- If you used **data**
 - Cite it with the **data**
- If you analyse **data**
 - Refer to the tool **data** with a persistent identifier
- List your data on your CV just as any other citation!

**Make your data FAIR
... and CARE for your
data**

**Follow
scientific
practices**

