

# CLARIN: AN OVERVIEW

CLARIN AND LIBRARIES WORKSHOP  
ROYAL LIBRARY OF THE NETHERLANDS, THE HAGUE  
2022, MAY 9TH & 10TH

Andreas Witt, National Coordinator CLARIN-D

Leibniz Institute for the German Language  
Mannheim  
Mannheim University  
(& Heidelberg University & University of Cologne)

---

# Main Goals of CLARIN

- Construction of an integrated, interoperable, and scalable research infrastructure via a network of centers
- Deployment of linguistic data, tools, and services
- Target group: researchers from the social sciences and humanities

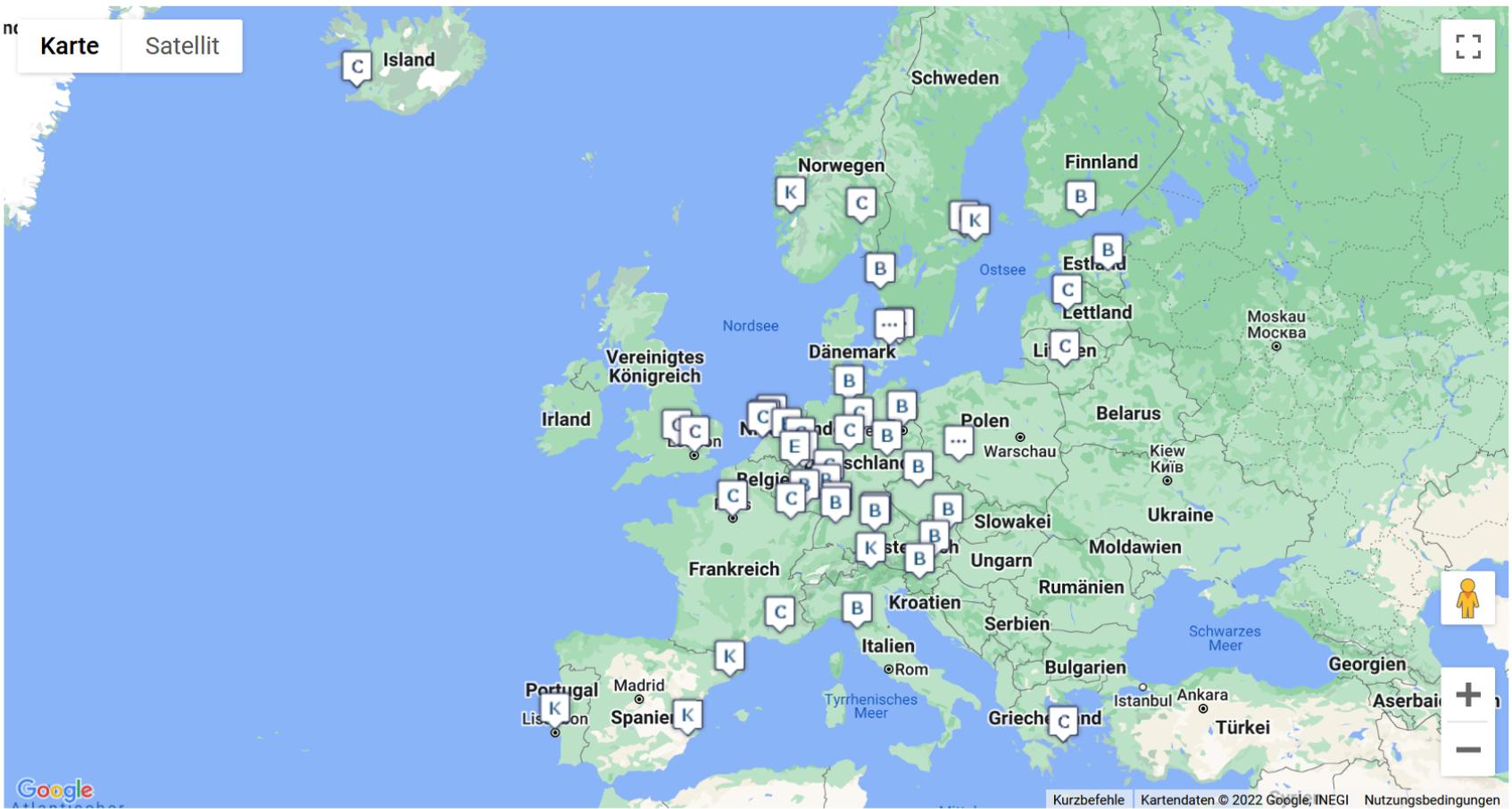
# Central structures of CLARIN

- CLARIN Centres: hosting the infrastructure
  - Central services (PID, search tools, ...)
  - Hosting data
  - Training and dissemination
  - Supporting community

# CLARIN

Google Kalender - Woche vom x DB Startseite - ICE Portal x cv\_mb\_05\_21 x Overview CLARIN Centres | CLA x +

https://www.clarin.eu/content/overview-clarin-centres



# CLARIN - organizational structure

- Legal status
    - ERIC: European Research Infrastructure Consortium
    - The Members of an ERIC are countries
  - Governance
    - General Assembly (representatives from the ministries of the member countries)
    - Scientific Advisory Board
    - Board of Directors
    - Several committees
-

# CLARIN - Government structure - BoD



## Board of Directors

The Board of Directors (*BoD*) is responsible for the day-to-day management of CLARIN *ERIC* and for the implementation of the strategies and policies set out by the General Assembly.



Franciska de Jong

Executive Director



Dieter Van Uytvanck

Vice Executive Director and Technical Director



Francesca Frontini

Member of the Board of Directors



Antal van den Bosch

Member of the Board of Directors

# CLARIN in Europe / Members

Members	National Consortium (NC)	Leading NC Partner	National Coordinator
Austria	<a href="#"><u>CLARIAH-AT</u></a>	ACDH-CH	Karlheinz Mörth
Belgium	<a href="#"><u>CLARIN BE</u></a>	Instituut voor de Nederlandse Taal (INT)	Vincent Vandeghinste
Bulgaria	<a href="#"><u>CLaDA-BG</u></a>	Bulgarian Academy of Sciences	Kiril Simov
Croatia	<a href="#"><u>HR-CLARIN</u></a>	University of Zagreb	Marko Tadić
Cyprus	CLARIN-CY	Digital Heritage Research Lab (Cyprus University of Technology)	Marinos Ioannides
Czech Republic	<a href="#"><u>LINDAT/CLARIAH-CZ</u></a>	Charles University Prague	Eva Hajičová
Denmark	<a href="#"><u>CLARIN-DK</u></a>	University of Copenhagen Center of	Costanza Navarretta

# CLARIN in Europe / Members

Members	National Consortium (NC)	Leading NC Partner	National Coordinator
Finland	<a href="#"><u>FIN-CLARIN</u></a>	University of Helsinki	Krister Lindén
Germany	<a href="#"><u>CLARIN-D</u></a>	Institut fuer Deutsche Sprache, University of Mannheim	Andreas Witt
Greece	<a href="#"><u>clarin:el</u></a>	ILSP-ATHENA Research Center	Stelios Piperidis
Hungary	<a href="#"><u>HunCLARIN</u></a>	Research Institute for Linguistics, Hungarian Academy of Sciences	Tamás Váradi
Iceland	<a href="#"><u>CLARIN Iceland</u></a>	The Árni Magnússon Institute for Icelandic Studies	Starkaður Barkarson

# CLARIN in Europe / Members

Members	National Consortium (NC)	Leading NC Partner	National Coordinator
Latvia	<a href="#"><u>CLARIN-LV</u></a>	Institute of Mathematics and Computer Science, University of Latvia	Inguna Skadiņa
Lithuania	<a href="#"><u>CLARIN-LT</u></a>	Vytautas Magnus University	Jurgita Vaičenonienė
The Netherlands	<a href="#"><u>CLARIAH-NL</u></a>	Utrecht University	Jan Odijk
Norway	<a href="#"><u>CLARINO</u></a>	University of Bergen	Koenraad De Smedt
Poland	<a href="#"><u>CLARIN PL</u></a>	Wroclaw University of Technology	Maciej Piasecki
Portugal	<a href="#"><u>PORTULAN CLARIN</u></a>	University of Lisbon	António Branco
Slovenia	<a href="#"><u>CLARIN.SI</u></a>	Jožef Stefan Institute	Tomaž Erjavec <sup>9</sup>
Sweden	<a href="#"><u>SWE CLARIN</u></a>	Språkbanken	Lars Borin

# CLARIN - Observers and 3<sup>rd</sup> party partners

<b>Observer</b>	<b>National Consortium (NC)</b>	<b>Leading Partner NC</b>	<b>National Coordinator</b>
South Africa	<a href="#"><u>SADiLaR</u></a>	North-West University	Langa Khumalo
United Kingdom	<a href="#"><u>CLARIN-UK</u></a>	Oxford University	Martin Wynne
<b>Third Party</b>	<b>National Consortium (NC)</b>	<b>Leading Partner NC</b>	<b>Coordinator</b>
Carnegie Mellon University (USA)	<a href="#"><u>TalkBank</u></a>	Carnegie Mellon University (USA)	Brian MacWhinney



CMU Talkbank

The Executive Director of CLARIN ERIC hereby acknowledges that the aforementioned institution has successfully completed the CLARIN Centre Assessment Procedure and is officially recognized as a

**CLARIN B Centre**

The aforementioned institution has permission to use the following logo in its official communications:



Franciska de Jong  
 Executive Director  
 CLARIN ERIC  
 Issued on: 12 October 2021  
 Valid till: 12 October 2024

Display K-centres

<b>Full name</b>	IMPACT centre of competence - CLARIN K-centre in digitisation
<b>Short name</b>	IMPACT-CKC
<b>URL</b>	<a href="https://www.digitisation.eu">https://www.digitisation.eu</a>
<b>Hosted by</b>	(1) Universidad de Alicante (UA), Alicante, Spain (2) Instituut voor de Nederlandse Taal (INT), Leiden, The Netherlands (3) Koninklijke Bibliotheek (KB), The Hague, The Netherlands (4) Poznan Supercomputing and Networking Center (PSNC), Poznan, Poland (5) The British Library (BL), London, United Kingdom (6) Bibliothèque nationale de France (BnF), Paris, France (7) Universidad de Guadalajara Mexico (UgG), Guadalajara, Mexico
<b>City of main hub</b>	Alicante
<b>Country of main hub</b>	ES
<b>Area of competence</b>	IMPACT-CKC (IMPACT centre of

CLARIN VLO
vlo.clarin.eu/jsessionId=5CC33DB7A06082A43F4318A9A4CFA25A?0

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

Language ⌵

Collection ⌵

Resource type ⌵

Modality ⌵

Format ⌵

Keyword ⌵

⌵

Availability ⌵

Search options ⌵

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

## WordNet

⊕ plWordNet is a lexico-semantic network which reflects the lexical system of the Polish language. There are at present ca. 144,000 nouns, verbs and adjectives in plWordNet, ca. 203,000 word senses and ca. 500,000 relations. It is already the second-largest wordnet in the world, and it keeps growing.

---

## Finnish WordNet

⊕ The language resource is available in Kielipankki - the Language Bank of Finland at <http://urn.fi/urn:nbn:fi:lb-2016042205>; download: <http://urn.fi/urn:nbn:fi:lb-2014052713>. The Finnish WordNet is a lexical database for Finnish. It is a part of the FIN-CLARIN infrastructure project. FinnWordNet is lic...

---

## Czech WordNet

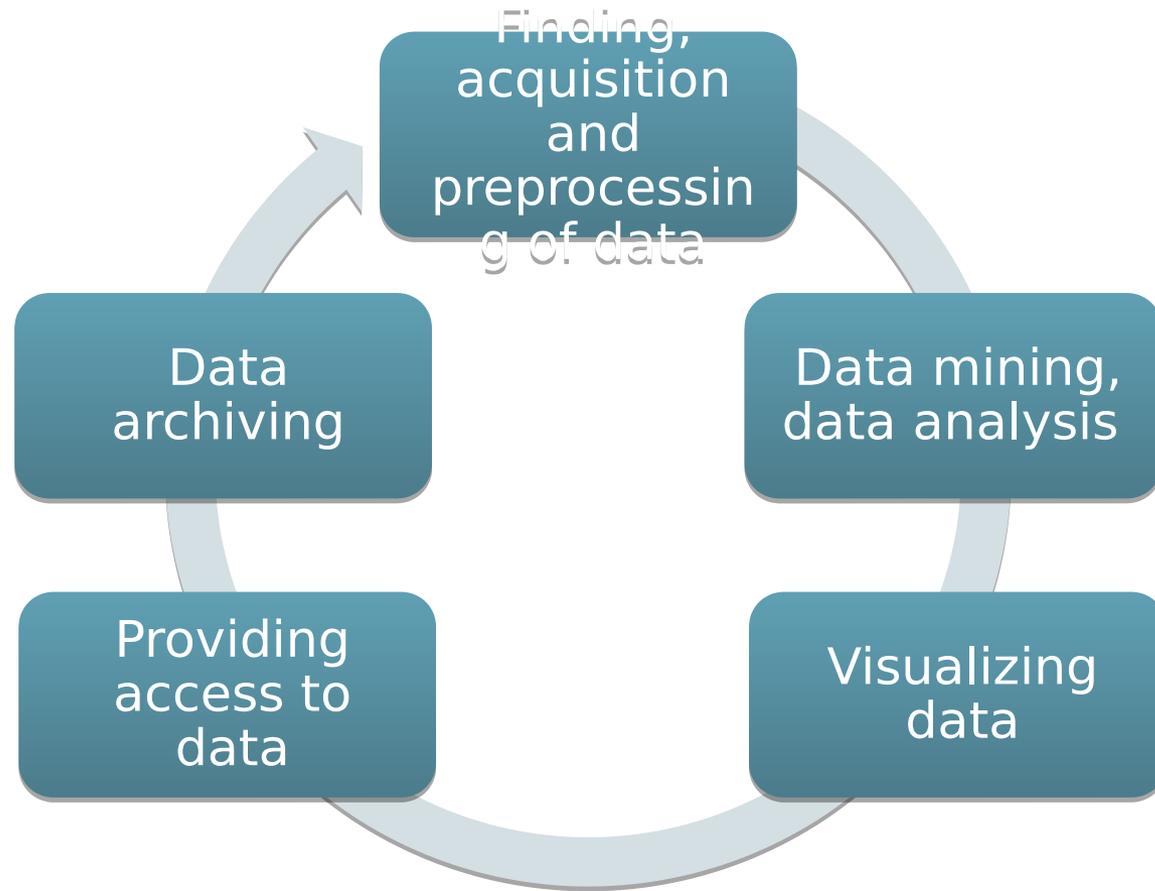
⊕ Multilingual Lexicons; The Czech WordNet was developed by the Centre of Natural Language Processing at the Faculty of Informatics, Masaryk University, Czech Republic. The Czech WordNet captures nouns, verbs, adjectives, and partly adverbs, and contains 28,201 word senses (synsets). Every synset encodes the equivalenc...

---

## Bulgarian WordNet

⊕ Multilingual Lexicons; The Bulgarian WordNet is a network of lexical-semantic

# Supporting five phases of e-Humanities projects



# Referencing resources

- Resources referenced via persistent identifiers (PIDs)
- PIDs resolve to a repository/archive
- If repository moves, resolvers change
- No requirement for "last visited" or similar

---

<sup>7</sup>All the sentence materials, instructions and a digital version of the questionnaire are freely available online: <http://hdl.handle.net/11022/0000-0000-1CCA-E> (Experiment 1), <http://hdl.handle.net/11022/0000-0000-1CCB-D> (Experiment 2), and <http://hdl.handle.net/11022/0000-0000-1CCC-C> (Experiment 3).

# Searching in resources



Zadejte slovo nebo slovní spojení

nerozlišovat velikost písmen  vyhledávat slovní tvary zadaného lemmatu

### Korpusový vyhledávač konText

Aplikace **konText** představuje základní vyhledávací program pro práci s korpusy projektu ČNK (nahrazuje tak původní NoSKE). Umožňuje tvorbu a vyhodnocení jednoduchých i komplexních dotazů, zobrazení formou konkordančních řádků, výpočtení frekvenční distribuce, asociálních měr kolokací a další práci s jazykovými daty. Přehledné jsou všechny funkce konTextu popsány v [manuálu](#).

**Jak to vlastně je?**  
 Jak mluvíme? Vlastně jako ani moc nevíme? Podívejme se na časté výpůjčky slovíčko vlastně do korpusu mluveného jazyka a srovnáme, jak se například liší mluva mladších a starších. Více se dozvíte v [manuálu](#) práce s korpusem.

V korpusu ORAL2008 vyberte typ dotazu *slovní tvar*, zadejte:

---

### Poradna a fórum uživatelů

V rámci [Fóra uživatelů](#), jehož členem se automaticky stává každý registrovaný uživatel ČNK, můžete klást dotazy ohledně použití našich korpusů a nástrojů, sdílet své zkušenosti a diskutovat s ostatními uživateli.

### Ústav Českého národního korpusu

Český národní korpus je akademický projekt založený v roce 1993 Ústavem Českého národního korpusu. Jeho cílem je mapovat vývoj češtiny (psané i mluvené) ve všech jejích složkách. Korpusy ČNK jsou po [bezplatné registraci](#) otevřeny všem uživatelům, kteří touží vědět, jak se čeština používá - [více informací](#).

Jazykový korpus je rozsáhlý soubor autentických textů, které jsou pomocí speciálních nástrojů vyhledávány jazykové jevy (např. slova a fráze) v reálných datech podloženy jazykový výzkum v rozsahu, který je pro výzkum nezbytný.

---

### Dostupné korpusy

Projekt Český národní korpus nabízí uživatelům mnoho jazykových typů výzkumů. K dispozici jsou korpusy psaného jazyka, mluveného jazyka, paralelní korpus [InterCorp](#) s více než 30 jazyky a specializovaných. Jejich přehled s podrobnějšími informacemi najdete [zde](#).

**DTA**
Suchen
Texte Projekt Dokumentation Impressum

in der Titelzisten im Korpus in der Dokumentation

## Ranke, Leopold von: Die römischen Päpste. Bd. 2. Berlin, 1836.

Bild: 0561:549 << vorherige Seite
nächste Seite >>

1:1 fit 71%
519

Kaiser Ferdinands II. im Jahre 1629.

bündnerischen Päffen bis an den Comersee, eingenommen: 35000 Mann stark flog alsdann dieses Heer längs der Adda und dem Oglio hinab. Noch einmal ward der Herzog von Mantua aufgefordert sich zu unterwerfen. Er erklärte, er stehet im Schutze des Königs von Frankreich, mit diesem müsse man unterhandeln. Indem nun die Deutschen sich gegen Mantua, die Spanier sich gegen Montserrat bewegten, erschienen auch die Franzosen zum zweiten Male. Sie machten auch dieß Mal Fortschritte; sie nahmen Saluzzo, Pinerolo: aber in der Hauptsache richteten sie nichts aus; nicht einmal den Herzog von Savoyen vermochten sie auf seine zu ihrem Willen zu nöthigen. Die Spanier begannen Cafale, die Deutschen nach kurzem Stillstand Mantua zu belagern: \*) sie hatten bei weitem das Uebergewicht.

Kein Wunder, wenn in dieser Lage der Dinge jetzt in Wien selbst Erinnerungen an die alte kaiserliche Hoheit laut wurden.

„Man werde den Italienern zeigen, daß es noch einen Kaiser gebe, man werde Rechnung mit ihnen halten.“

„Man werde den Italienern zeigen, daß es noch einen Kaiser gebe, man werde Rechnung mit ihnen halten.“

Befonders hatte sich Venedig den Haß des Hauses Oestreich zugezogen. Man urtheilte zu Wien, daß wenn Mantua einmal gefallen, auch die Terra ferma von Venedig nicht mehr widerstehn könne. In ein paar Monaten müßte man sie haben, dann könne man die kaiserlichen Lehen zurückfordern. Der spanische Gefandte ging noch weiter. Er verglich die spanisch-österreichische Macht mit der römischen.

\*) Das alte Buch dell'istoria di Pietro Giov. Copriata erörtert die einzelnen Momente dieser Ereigniß.

Suomi Ruotsi Muut kielet Rinnakkaistekstit

Yksinkertainen Laajennettu Edistynyt

Etsi 
Etsi myös  alkuosa  loppuosa ja  samaista-pien-ja suuraakoset

Konkordanssi: osuvia sivulla: 25
järjestä korpuksen sisällä: järjestämätön
Tilastoja: laske tilastot tämän perusteella: sana

# Federated Content Search (FCS)

- Distributed search in a common interface
- No central index
  - Respecting IPR
  - Using existing search environments
- Webservice based
  - Rest-based webservices
  - Chains of tools with integrated AAI (no individual login required)

The screenshot displays the CLARIN-D Federated Content Search interface. At the top, the logo and title 'CLARIN-D FEDERATED CONTENT SEARCH' are visible. Below the title is a search bar containing the word 'Bibliothek' and a 'Search' button. Navigation links for 'About', 'Search options', 'Search results', and 'Help' are provided. A toolbar includes 'Clear', 'Export to Personal Workspace', and 'Download' options.

The main content area shows search results from the 'DTA Corpus, Berlin-Brandenburg Academy of Sciences and Humanities'. The results are presented in a table-like format with columns for text snippets and 'Bibliothek' labels. A red box highlights a specific snippet: 'Sie wurden an hohe Potentaten verschenkt, sie finden sich jetzt nur in den Königlichen Bibliotheken von Schweden Gustav Adolf führte sein Exemplar im Felde immer mit sich --, England, Dänemark.'

Below this, there are more search results from the 'TIGER, IMS, Universität Stuttgart' corpus, including a snippet about a photo at the Uni-Eingang and another about a person's access to a computer.

At the bottom, the copyright notice 'Copyright © 2013 Clarin-D' is displayed.

# CLARIN and Standards



# CLARIN and standards

Why are standards important for CLARIN?

- The project partners and centers is heterogeneous in both topics and organizational structure
  - The resources available are heterogeneous regarding topics, technical formats, and copyright
  - Standards may help to establish interchange
-

# Standards and CLARIN

- In CLARIN, everyone agrees that the use of standards is desirable to interoperate with each other
  - But: many researchers still lack knowledge to use 'the right standards'
  - Even if information about a specific standard is available, information about its relations to other standards is not
-

# The CLARIN Standards Committee

## The CLARIN Standards Committee (CSC)

- Members can be anyone involved in CLARIN
- Chair: Piotr Bański, Germany
- Vice-Chair: Karlheinz Mörth, Austria
- See

<https://www.clarin.eu/governance/standards-committee>

for a list of current members

---

# Legal & ethical issues in CLARIN



# Legal frameworks concerning language resources

- Copyright
- *sui generis* database right
- Data protection

# CLIC - CLARIN Legal Issues Committee

- Legal experts from every national consortium
- **Chair:** Pawel Kamocki - here in the room
- Regular meetings
- reports to the Board of Directors
- coordination of CLARIN ERIC's lobbying efforts at the EU level
- evaluation of relevant bills at the EU level

# CLARIN's 'legal offer' for researchers

- CLARIN Legal Infrastructure & standard licence terms
- CLARIN Legal Information Platform
- CLARIN Public License Selector & Calculator
- CLARIN-D Legal Helpdesk
- Lobbying & Networking

# Clarin Annual Conferences



# Clarín Annual Conferences

- CLARIN 2021, 27-29 September 2021, Virtual event
- CLARIN 2020, 5-7 October 2020, Virtual Event
- CLARIN 2019, 30 September - 2 October 2019, Leipzig, Germany
- CLARIN 2018, 8-10 October 2018, Pisa, Italy
- CLARIN 2017, 18-20 September 2017 , Budapest, Hungary
- CLARIN 2016, 26-28 September 2016, Aix-en-Provence, France
- CLARIN 2015, 15-17 October 2015, Wrocław, Poland
- CLARIN 2014, 23-25 October 2014, Soesterberg, The Netherlands
- CLARIN 2013, 21-22 October 2013, Prague, Czech Republic
- CLARIN 2012, 24-27 October 2012, Sofia, Bulgaria



Home

# CLARIN Annual Conference 2022

10 October 2022 - 12 October 2022



St. Vitus Cathedral, Prague Castle, Prague, Czechia

## TABLE OF CONTENTS

- [About the CLARIN Annual Conference](#)
- [Registration](#)
- [Call for Abstracts](#)
- [Open Calls](#)
- [Keynotes](#)

## About the CLARIN Annual Conference

The CLARIN Annual Conference is the main annual event for those working on the construction and operation of CLARIN across Europe, as well as for representatives of the communities of use in the humanities, and social sciences.

*In 2020 and 2021, the CLARIN Annual Conference took place virtually. In 2022, we hope to at least partially return to the traditional format, with Prague, Czechia, as the venue. The event will be organised by CLARIN ERIC in collaboration with LINDAT/CLARIAH-CZ.*

CLARIN2022 is organised for the wider humanities and social sciences communities in order to exchange

# CLARIN: AN OVERVIEW

Andreas Witt

Leibniz Institute for the German Language  
Mannheim

Mannheim University

(& Heidelberg University & University of Cologne)

---