



Common Language Resources and Technology Infrastructure



CLARIN is the Common Language Resources and Technology Infrastructure in Europe. Its main aim is to provide scholars in the humanities and social sciences with easy access to digital language data (in written, spoken, video or multimodal form) and to advanced tools for finding, exploring, using, annotating, analysing or combining such data, no matter where it is located. In more detail, CLARIN offers:

- Comprehensive services to the humanities disciplines with respect to language resources and technology.
- A persistent and stable infrastructure that researchers can rely on decades to come.
- Technology for overcoming the many barriers created by institutional, structural and semantic interoperability problems and fragmenting the resources and tools landscape.
- Tools and resources that will be interoperable across languages and domains, thus addressing the issue of preserving and supporting the multilingual and multicultural European heritage.
- Comprehensive training and education programs that include university education in the different member states.
- Improvement and extension of web-based collaborations, i.e. creating virtual working groups breaking the discipline boundaries.
- Development or improvement of standards for language resource maintenance.

CLARIN is building a networked federation of language data repositories, service centres and centres of expertise, with single sign-on access for all members of the academic community in all participating countries. Tools and data from the different centres will be interoperable, so that data collections can be combined and tools from different sources can be chained to perform complex operations to support researchers in their work.

The Scientific Challenge

CLARIN centres are working together with a group of central developers to connect language resources and tools so researchers in the humanities and social sciences can easily access and make use of these valuable resources. CLARIN wants to make it possible for “homeless” researchers and citizen scientists to deposit their resources into a good data repository with long-term preservation.

Who benefits and how?

Data stored in repositories is well-preserved in the long term, safe backups are in place, and large research data sets can be safely and conveniently shared. Not having to worry about these issues definitely makes a huge difference for the CLARIN community – scholars in the humanities and social sciences sectors.



Preliminary Results

CLARIN has been one of EUDAT's core communities since 2011 and the service that it has been most involved with, up to now, is [B2SAFE](#). Various CLARIN centres are using it to perform safe replication of the language data they are hosting. The University of Tübingen (Eberhard Karls Universität Tübingen), the LINDAT/CLARIN Centre for Language Research Infrastructure in the Czech Republic (usually known as LINDAT/CLARIN) and the Max Planck Institute for Psycholinguistics in the Netherlands are all using B2SAFE. There are a further eight centres ready to use B2SAFE and tailored uptake plans are under development to be deployed over the coming months.

CLARIN will harvest the [B2SHARE](#) metadata related to language material and make it accessible via their search portal, the [Virtual Language Observatory](#). Additionally, EUDAT's [B2DROP](#) service has been tested and is used for internal data exchange and sharing. CLARIN community metadata has been integrated into [B2FIND](#), EUDAT's metadata portal.

Contacts

- Dieter Van Uytvanck, CLARIN ERIC, [dieter\(at\)clarin.eu](mailto:dieter(at)clarin.eu)

Further Information

[Download the presentation for more information on the EUDAT and CLARIN collaboration](#)

Read our [interview](#) with Dieter Van Uytvanck, the Technical Director for CLARIN ERIC, about CLARIN's ongoing work integrating its infrastructure with the EUDAT common data services. Dieter is based at Utrecht University where he is responsible for the construction and maintenance of CLARIN's technical infrastructure.

Category:

[Core Communities](#)

[Read more](#)