
# VALUE PROPOSITION

CLARIN-PLUS receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no. 676529.

www.clarin.eu

# Table of contents

# 1 Value proposition at a glance

CLARIN is a networked federation of language data repositories, service centres and centres of expertise. CLARIN makes digital language resources available to scholars, researchers, students and citizen-scientists from all disciplines, especially in the humanities and social sciences, through single sign-on access. CLARIN offers long-term solutions and technology services for deploying, connecting, analyzing and sustaining digital language data and tools. CLARIN supports scholars who want to engage in cutting edge data-driven research, contributing to a truly multilingual European Research Area.

### Mission

Create and maintain an infrastructure to support the sharing, use and sustainability of language data and tools for research in the humanities and social sciences.

### Vision

All digital language resources and tools from all over Europe and beyond are accessible through a single sign-on online environment for the support of researchers in the humanities and social sciences.

### Disciplines

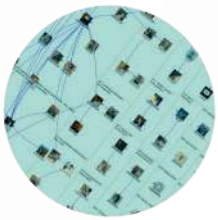
CLARIN stimulates the reuse and repurposing of available language data, thereby enabling scholars in the (digital) humanities and the social sciences to open new research avenues within and across disciplines that address one or more of the multiple societal roles of language. Why is language so important? It is a carrier of cultural content and information, both synchronically and diachronically, but it also plays a role as the reflection of scientific and societal knowledge, as an instrument for human communication, as one of the central components of the identity of individuals, groups, cultures or nations, as an instrument for human cognition and expression, and as an object of study or preservation.

## Open Science

More generally CLARIN does not see itself as a stand-alone facility, but rather as a player in making the vision that is underlying the emerging European policies towards Open Science a reality, interconnecting researchers across national and discipline borders by offering seamless access to data and services in line with the FAIR data principles.

## Stakeholders

The construction and operation of an infrastructure involves many different stakeholders, each with their own interests and expectations, ranging from individual researchers, research institutions, data archives, infrastructure service providers, funding bodies and governments, and to sectors that are not primarily academic, such as the data industry and GLAM (galleries, libraries, archives, and museums) – just to mention a few.



"CLARIN offers infrastructural support for the study and use of language resources as social and cultural data"

*Franciska de Jong, CLARIN ERIC*

# 2 Value proposition for researchers

## 2.1 Discovering and depositing resources

Researchers can search for language resources via metadata – in the CLARIN "catalogue", the Virtual Language Observatory (VLO), or search in the data itself (Content Search). The VLO contains references to more than 800,000 resources the majority of which are hosted at CLARIN centres, but the VLO also contains references to other relevant resource collections.

**Metadata search:** www.clarin.eu/vlo

**Content search:** www.clarin.eu/contentsearch

As CLARIN ERIC has now many different countries as members, the VLO covers many languages, both national and regional languages, as well as languages studied in those countries.

The advantage of the VLO is faster identification of relevant resources, allowing researchers to re-use resources that already exist, rather than having to produce their own. Additionally, the VLO allows users who create or collect their own datasets to make them better visible to others through publication of the metadata in the VLO.

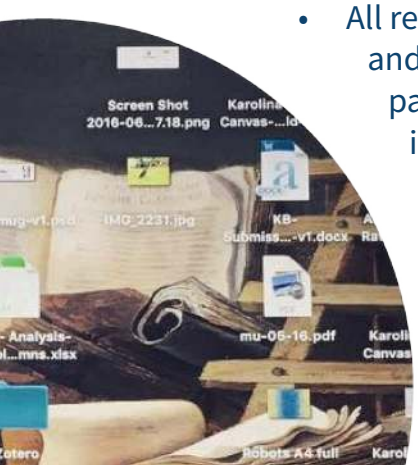
Without the VLO, researchers would have to use other catalogues, or general purpose search engines, often with less precise results – the VLO has a high number of query facets which can be used to guide the search and make the result more precise.

Researchers will also profit from the advantages of the cross border coordination which lies behind the VLO: it is easier not only to locate, but also to get access to CLARIN resources.

## Long-term preservation

One of the fundamental services of the CLARIN infrastructure is making sure that language resources can be archived and made available to the community in a safe and sustainable manner. To help researchers to store their resources (e.g. corpora, lexica, audio and video recordings, annotations, grammars, etc.) in a sustainable way, at least one CLARIN data centre in each country offers a **depositing service**. These centres are willing to store the resources in their repository and assist with the technical and organisational details. This has a wide range of advantages:

- Long-term archiving: In every CLARIN member country there are one or more CLARIN data centres that have committed themselves to offering a storage guarantee for a longer period of time.
- The resources can be cited easily and reliably as they have a persistent identifier.
- All resources and their metadata are equally accessible and searchable throughout the CLARIN infrastructure, irrespective of their physical location.
- Online searchable accessibility of all resources and their metadata throughout the CLARIN infrastructure, irrespective of their physical location.
- Online access to all resources and their metadata throughout the CLARIN infrastructure, irrespective of their physical location.
- All resources and their metadata are equally accessible and searchable 'in the cloud' throughout the CLARIN infrastructure.
- All resources and their metadata can be combined, analysed and enriched with various linguistic tools (e.g. automated part-of-speech tagging [https://www.clarin.eu/node/3441), information extraction, phonetic alignment (http://clarin.phonetik.uni-muenchen.de/BASWebServices ) or audio/video analysis (http://tla.mpi.nl/projects_info/avatech avatech-results/)], irrespective of the physical location of data and tools.



## Researchers are providers

Researchers are not just consumers of data and tools, but also providers in that they are encouraged to share their data and tools with others, if necessary in a protected way, so that others can build further on their results. This is supported by the availability of repositories (see above) and by facilitating data citation and licensing (see below).

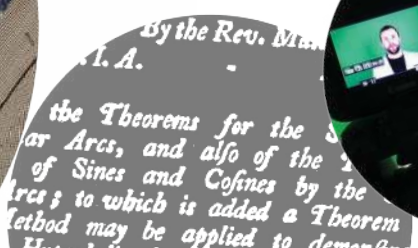
## Data citation

It is a major task to produce a corpus or a data collection. More and more, the scientific world recognises the value of such contributions, and mechanisms have been developed for data citation that encourage creators of corpora or other data collections to publish their data. CLARIN is a perfect platform for this type of publication and subsequent citation. This is because CLARIN offers a good search tool (VLO) for better publicity and better sharing, and because of the use of persistent identifiers to refer to them instead of notoriously unstable URLs.

## Licensing

CLARIN centres make data available through licensing and clear conditions for use. This involves CLARIN centres making deals with rights owners, signing Deposition License Agreements which include End User License Agreements, categorizing licenses in clearly marked license categories, and writing Terms of Service.

Guidance is offered to creators of data who want to select the most appropriate licensing conditions when publishing their data.



## 2.2 Advanced tools and computing facilities

CLARIN offers state-of-the-art tools and online services for many languages. These support researchers to annotate, analyse and publish their language data. Automatic analyses such as e.g. annotation can be performed faster and more easily on large amounts of data, and data and tools from different sites can be combined.
Examples of the functionality offered:

**Advanced analysis and visualizations for large data sets that may help gaining deeper insights, e.g.:**

- DiaCollo: collocation analysis in diachronic perspective (www.clarin.eu/showcase/diacollo)
- Stylo: state-of-the-art tool for stylometric analysis (http://clarin-pl.eu/en/services)

**Faster automated analysis, leading to more time for the actual research, e.g.:**

- WebMAUS: Automatic Segmentation and Labelling of Speech Signals over the Web (https://www.clarin.eu/showcase/webmaus-automatic-segmentation-and-labelling-speech-signals-over-web)
- AVAtech: audio and video recognizers (https://tla.mpi.nl/projects_info/avatech/avatech-results/)

**Reproducible scientific analysis flows, leading to more data sharing and better replicability of research results, e.g.:**

- Mind Repository: a platform for researchers to share their papers together with the data they have used and the scripts to analyse them (http://openscience.uni-leipzig.de/)
- Web service orchestration engines: platforms that allow researchers to combine different tools (possibly residing on different servers) in order to perform complex operations on their data (https://www.clarin.eu/content/web-services)

**Access to first-class corpora through specialized query interfaces, e.g.:**

- The Corpus of Contemporary Dutch (https://portal.clarin.inl.nl/)
- The International Computer Archive of Modern and Medieval English (ICAME) corpora (http://clu.uni.no/icame/clarin/)
- Data sets that support comparative research, such as parliamentary data, e.g. LinkedEP dataset (http://linkedpolitics.ops.few.vu.nl/web/html/home.html)

For large and computation-intensive tasks, CLARIN can connect scientists to highly ranked High Performance Computing (HPC) centres. Depending on the amount of computing resources needed, the researcher might need to enrol in a competitive call. In any case there are no costs for the use of the HPC facilities.

## 2.3 Federated login: easier access to more resources

CLARIN has established a Service Provider Federation, i.e. a trusted network of identity providers which offer "single sign-on". This means that researchers can login with their institutional credentials to get access to password-protected language resources and applications in other countries.

One advantage for the researchers is that they gain time and have the benefit of having to use only one access code. Another advantage is that they also get access to protected resources in other countries. Without this single sign-on, researchers would either have no access to otherwise valuable resources, or they would have to apply for accounts for each repository.

Statistics show that about 60 unique visitors per day use the federated login to access CLARIN resources[1].

You can find some of the online resources to which CLARIN gives easy access here:
https://www.clarin.eu/content/easy-access-protected-resources

[1] Measured using Piwik during the first half of 2016 at the CLARIN discovery service. As not all Service Providers are using this service and CLARIN respects users that opt-out of tracking it is an underestimation of the real number. These remarks also apply for all other usage statistics in this document. See CE-2015-0528 (https://www.clarin.eu/content/piwik-background-information) for details.

## 2.4 Access to expertise for researchers

Complementary to the access to data and tools CLARIN offers researchers access to expertise through its Knowledge Sharing Infrastructure. First of all, there are help desks and knowledge centres.

All CLARIN centres offering access to data and tools operate a helpdesk (in English) where (potential) users from all CLARIN countries can get information about the data and services offered, get help in using the services, and report problems.

Certified CLARIN knowledge centres (https://www.clarin.eu/content/knowledge-centres) offer access to expertise in specific areas, such as treebanking, speech analysis, audio-visual fieldwork, language learning, the Danish language, the languages of Spain, the languages of Sweden.

Specialized committees bring together expert knowledge about various topics, such as IPR and licenses (CLIC – the CLARIN Legal Issues Committee) and standards (STAC - the Standards Committee).

The CLARIN ERIC office operates a central helpdesk that will reply directly to requests for help or information, or channel the requests to the best experts.

## 2.5 Online tutorials etc.

The Knowledge Sharing Infrastructure also offers online tutorials and explanations of use cases in order to facilitate the uptake of new techniques by researchers. These materials are provided by experts from national CLARIN consortia, and wherever possible, offered at least in English or (in the case of movies) with English subtitles in order to facilitate cross-border access.

The use cases serve to demonstrate the successful application of digital methods to specific research questions, and to inspire researchers with similar or related questions.

## 2.6 Workshops and seminars, mobility grants

The final part of the knowledge sharing infrastructure consists of direct and physical meetings of various types.

Through an open call for workshop proposals members of the CLARIN community are invited to submit proposals for workshops on strategic priorities for CLARIN, or for the preparation of small development projects.

From time to time CLARIN ERIC organises workshops focused on specific topics where members of the CLARIN community get together and work on topics of common interest.

At the Annual CLARIN Conference infrastructure providers and users from all CLARIN countries exchange expertise and ideas.

A cross-border mobility scheme for short stays at CLARIN centres helps the exchange of knowledge and expertise between staff of different centres, or between staff from centres and researchers.

## 2.7 Support for data management plans

Data Management Plans (DMPs) are now becoming an indispensable and under many funding schemes even mandatory component of project plans that aim at collecting or creating new data, or at enriching existing data collections. The objective of DMPs is to secure the continued availability of project data beyond the duration of the project, thus enabling replication of results and advancing research by allowing researchers to build on each other's results.

Some national CLARIN teams have already elaborated instruments that help the researcher in creating Data Management Plans, e.g. DMPTY – A Wizard For Generating Data Management Plans (Trippel and Zinn, 2015; CLARIN-D).

In the context of the H2020 project PARTHENOS CLARIN, DARIAH and a number of other European infrastructure initiatives in the humanities and social sciences are making a joint effort to develop an easy to use template plus corresponding guidelines, to be used by researchers when preparing their project plans.

# 3 Value Proposition for institutions and institutes

## 3.1 Guaranteed access to CLARIN central services

CLARIN offers a wide range of central services from which all users and all CLARIN centres in CLARIN ERIC member countries and third parties with specific contractual agreements benefit.[2] These services are developed and offered in close collaboration between the national teams and CLARIN ERIC. Here we mention some of those services as illustration. More examples can be found in section 2.:

- The central Discovery Service with 24/7 availability (https://www.clarin.eu/content/clarin-discovery-service)
- As official EUDAT community, access to advanced EUDAT services: e.g. B2DROP (workspaces; http://eudat.eu/services/b2drop) and B2SAFE (safe replication; http://eudat.eu/services/b2safe)
- Services for creating, sharing and re-using rich metadata descriptions: Component Registry (creation and editing of metadata schemas), Concept Registry (registration and re-use of semantic definitions), OAI harvester (automated distribution of metadata files)
- The CLARIN centre registry, a central dashboard to register services and to connect them to automated checking and publication processes (monitoring, harvesting, content search, etc.)
- Advanced distributed user statistics, using Piwik.

Institutions and institutes are not only the homes of users of data and services, but they are also providers. Research institutes can share the results and spin-offs of their research through CLARIN and make them more visible, and institutions with an archiving function can use CLARIN to make their holdings, often created with public funding, more widely visible and usable.

[2] In accordance with what is specified in the statutes, Art. 18.

## 3.2 Participation in European projects

CLARIN ERIC is actively applying for participation in European projects when such participation supports CLARIN's general strategy and/or development. This means that joining CLARIN ERIC opens enhanced opportunities for participation in European funded projects, as staff of national CLARIN consortia may work on such projects on behalf of CLARIN ERIC.


In the period 2015-2017 CLARIN ERIC is coordinating the H2020 project CLARIN-PLUS, which aims at further consolidation and expansion of the CLARIN infrastructure, both technically and organisationally.

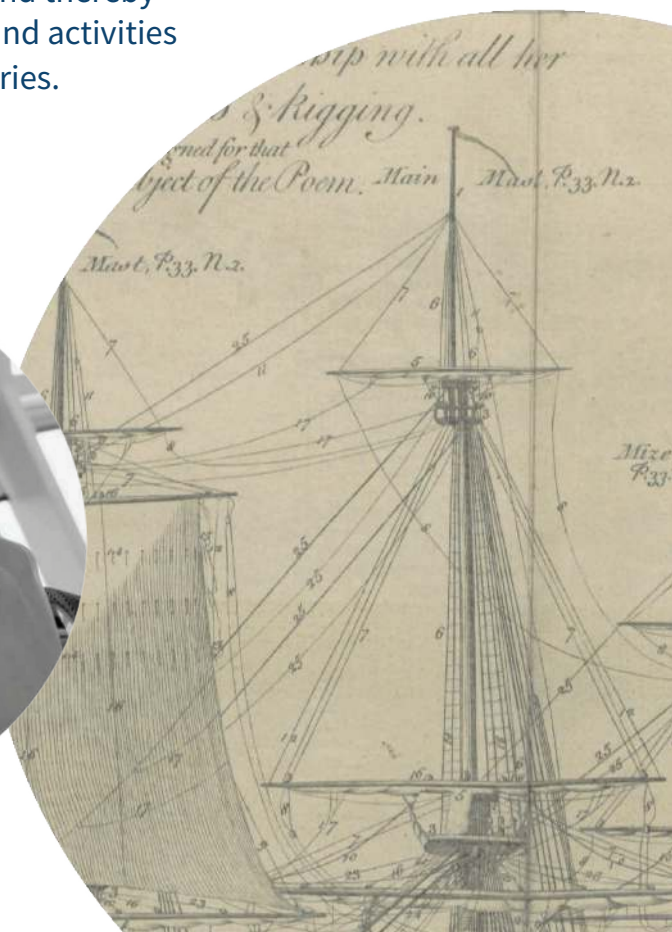
Additionally, CLARIN ERIC has been participating in four other European projects: PARTHENOS, EUDAT2020, Language Technology Observatory and EUROPEANA-DSI. At this moment staff from more than a dozen CLARIN ERIC Member or Observer countries is participating in EU projects on behalf of CLARIN ERIC for a broad variety of tasks in terms of topic and size. More information about ongoing EU projects with CLARIN ERIC involvement can be found on a special section of our web: https://www.clarin.eu/content/clarin-eu-projects.


## 3.3 Access to expertise for CLARIN institutions and institutes

To institutions planning to establish themselves as CLARIN technical centres or aiming to keep themselves up-to-date with recent developments we offer various types of support, including funding for travel costs involved.

- Best practice and other documentation on the creation and maintenance of infrastructure services is available, and will be continuously updated.
- Thematic workshops or tutorial sessions on specific infrastructure topics are organized to bring together centre staff from new and established centres
- A mobility scheme for short (typically up to 1 week) exchanges between new and established centres.

To all institutions and institutes looking for opportunities for collaboration with relevant parties across borders CLARIN offers access to expertise and experts, and thereby facilitates the creation of new projects and activities across the borders of regions and countries.





“Through a current search-engine researchers get access to valuable linguistic information. Everybody has access and can share information. Here interoperability is key; at a technical level we must all speak the same language.”

*Dieter Van Uytvanck, CLARIN ERIC*


# 4 Value Proposition for countries

In addition to the above-mentioned benefits for a country's researchers and institutions, the following national strategic advantages of joining CLARIN ERIC can be identified:

CLARIN membership may contribute to a better position of the national language(s) in the European and international context by making data more visible and more widely accessible.

Visibility of cultural content increases: As much of a country's cultural content is of linguistic nature or described by means of language, disclosing it through CLARIN will make it more visible, both for the research community and for the public at large. Conversely membership of CLARIN ERIC will give researchers in the country full access to the cultural heritage in other CLARIN countries.

As generic infrastructure services can be used across borders, CLARIN members can benefit from the fact that the costs of construction and operation of such services can be shared between members. Software of a generic nature can be shared between members, and in many cases the development cost of more specific software can be reduced by porting it between languages.

Not only will the development cost be smaller, the access to CLARIN resources (data, tools and methods) will also lead to more advanced research and open new research avenues across borders and disciplines, as researchers can build on each other's results on a European scale.

CLARIN members have direct influence on the decision-making about all aspects of the infrastructure, ranging from construction and operation, to longer term evolution and strategic priorities, as voting member in the General Assembly, the National Coordinators' Forum, etc., as well as through participation in the various committees.

More details about the governance of CLARIN ERIC can be found in the Statutes, see http://hdl.handle.net/1839/00-DOCS.CLARIN.EU-75.




**Access to expertise for new CLARIN countries:**

For emerging national consortia in countries that have just joined or are preparing themselves for joining CLARIN, CLARIN offers workshops on how to set up the CLARIN infrastructure at the national level. Four workshops for new and future members are being held 2015-2017, and the information materials produced are available on the CLARIN website.

Additionally, a series of best practice papers has been prepared on the following themes:

- Building a national consortium
- Living up to the criteria for CLARIN centres
- Building the Knowledge Sharing Infrastructure at the national level
- Cost estimates for the construction and operation of the national infrastructure
- Professional profiles for typical CLARIN infrastructure functions
- IPR and ethical issues
- Interoperability and standards

# 5 Values for Europe

### Excellence

CLARIN's role is not just to offer infrastructure services per se, but also to contribute to the further development of the European Research Area. In this context CLARIN has a number of things to offer to support cutting edge research and excellence in European research in a broad variety of disciplines where language is a significant data type, such as linguistics, language teaching, literary research, history, political science, cultural heritage, social sciences etc.

### Cross-border, cross-language research

CLARIN offers the same access to data and services to all users. This is a necessity for joint research and for the replication of research results, which has become a very important factor in many disciplines. At the same time (and maybe even more importantly) CLARIN is also offering access to similar data collections and services to the European humanities and social sciences research community at large. Examples of such data types are newspaper collections, corpora of parliamentary records and social media data sets. In addition to offering harmonized access these collections CLARIN has taken up the coordinating role of creating overviews of such comparable datasets and of bringing together researchers from various disciplines to explore the options for joint cross-border and cross-language research, and exchange of expertise. This will stimulate the collaboration for communities with a research agenda that requires the exploration of language data from a comparative perspective. As language is a data type that captures linguistic, cultural and social phenomena, the potential for multidisciplinary research that addresses Europe's societal challenges is enormous.

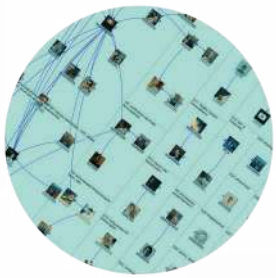
## Open Science

CLARIN is not a stand-alone facility, but is well-embedded in the European research infrastructure landscape at large, and as such fully committed to the European Open Science Policies, including re-use of data.
As mentioned in 4.1.7 a focus on data management is inherent in CLARIN's data preservation activities. Through its focus on language it offers opportunities for cross-discipline collaboration between areas where language plays a role, both within the humanities and social sciences and in other disciplines.

Various instruments will be used to support this, including:

- Workshops on typical cross-border, cross-language or cross-discipline topics in order to ensure a continuous flow of knowledge and expertise between different communities

- Initiatives for or participation in EU projects with a strong language dimension that bring together researchers from national CLARIN consortia, thereby providing excellent input and expertise for these projects.





CLARIN in three words: "Internationality, openness and user-friendliness!"

*Tommi Jantunen, University of Jyväskylä*

## References


CLARIN ERIC Statutes. Official Journal of the European Union L64, 2012, pp.13-28. (http://hdl.handle.net/1839/00-DOCS.CLARIN.EU-75)

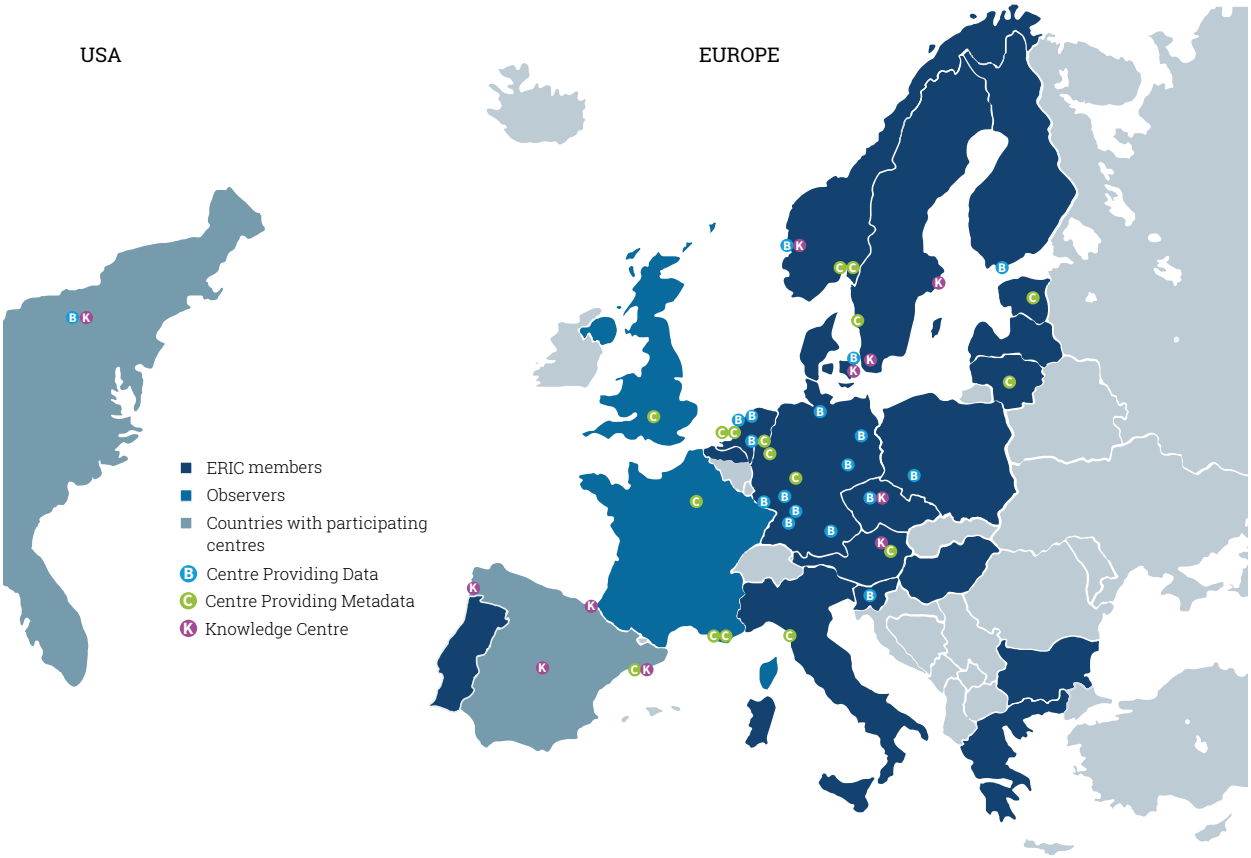
The European Strategy Forum on Research Infrastructures, Strategy Report of Research Infrastructures, 2016. (section on CLARIN ERIC, p.82.) (http://www.esfri.eu/roadmap-2016)

Trippel, Thorsten und Claus Zinn (2015). DMPTY - A Wizard For Generating Data Management Plans, CLARIN Annual Conference, Wroclaw (Poland), 2015. (http://www.ep.liu.se/ecp/article.asp?issue=123&article=006&volume=)

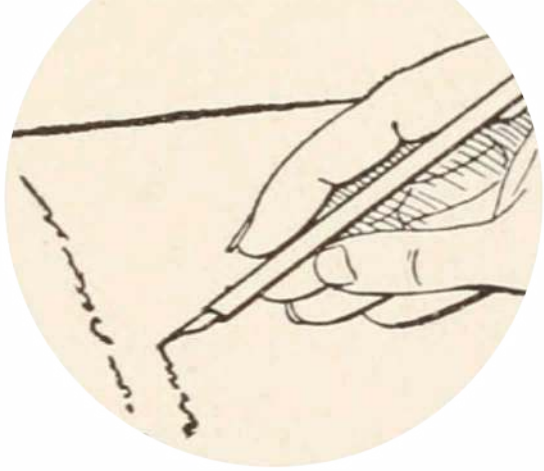

*Throughout the document many links are made to the CLARIN ERIC website, www.clarin.eu, and other websites where documentation can be found. These links are not repeated here.*

# CLARIN Centres

*Map as of July 2017*




# Notes



## Colophon

The content of this document is based on one of the deliverables of the H2020 project CLARIN-PLUS:

> D5.4 CLARIN Value Proposition, CE-2016-0847, August 2016 (https://office.clarin.eu/v/CE-2016-0847-CLARINPLUS-D5_4.pdf); authors: Bente Maegaard, Dieter Van Uytvanck, Steven Krauwer.

**Editing & coordination:**
CLARIN ERIC

**Design:**
Karolina Badzmierowska

**Online version:**
https://www.clarin.eu/value-proposition

**Publication number:**
CLARIN-CE-2017-1093-P001
September 2017

**Contact:**
CLARIN ERIC
c/o Utrecht University
Drift 10, 3512 BS Utrecht
The Netherlands

**www.clarin.eu**

CLARIN ERIC
Common Language Resources and Technology Infrastructure