Common Language Resources and Technology Infrastructure

CLARIN

A European Research Infrastructure

Towards a persistent and stable research infrastructure offering easy access to interoperable language resources and technology for the Humanities and Social Sciences

Language is at the heart of many disciplines in the Humanities and Social Sciences as
- an instrument for communication,
- a means to represent information and knowledge, and
- an object of study in itself.

Many collections of digital content exist, and technologies and tools have been developed to exploit, retrieve, explore, enhance and study these data. However, these resources and tools are
- geographically scattered,
- technically fragmented, and
- difficult to find, access and use.

Therefore, CLARIN is
- creating and maintaining a sustainable infrastructure
- to enable high quality, innovative digital research in the Humanities and Social Sciences.

CLARIN is based on four pillars:
- digital language content and tools,
- a technical infrastructure to provide access for the users,
- a knowledge sharing infrastructure to provide support,
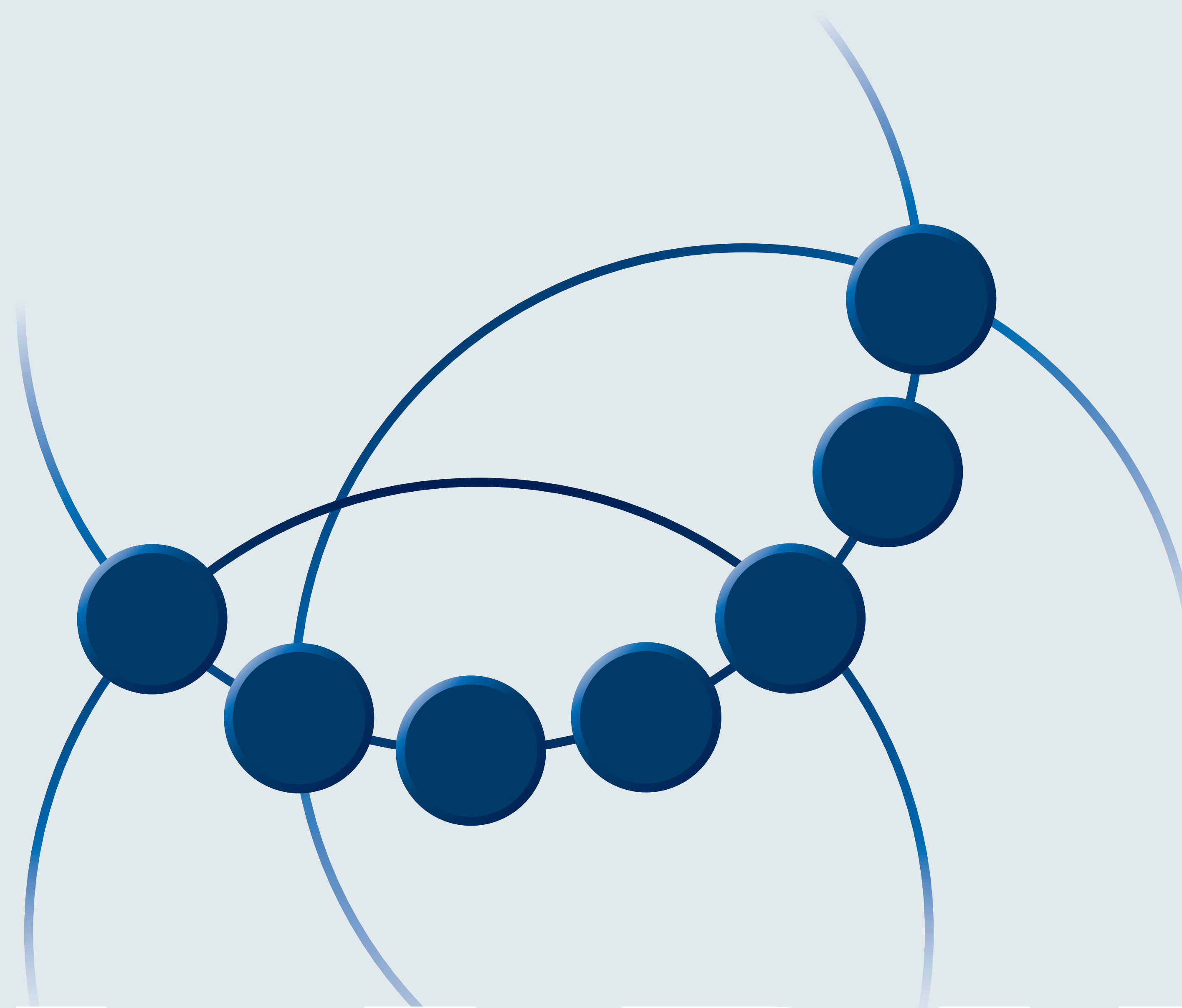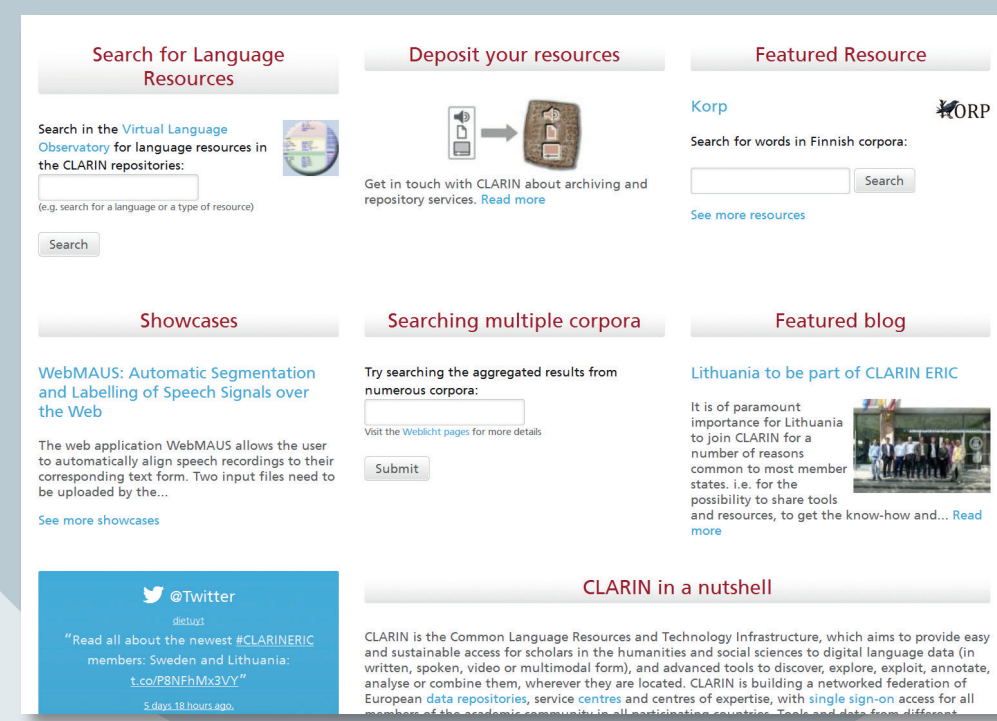- an efficient and sustainable organisational framework.

Important CLARIN principles are:
- interoperability of data and tools,
- stability and availability of services,
- long term accessibility,
- adaptability and extensibility,
- partnerships with other European infrastructures,
- embedding in the international research infrastructure landscape, and
- commitment to open access.

Governance and funding:
- CLARIN ERIC (European Research Infrastructure Consortium) was set up February 2012;
- founding coalition of eight countries and one intergovernmental organisation;
- more countries have joined and are joining;
- hosted by the Netherlands;
- jointly funded by the member countries & organizations.

at.clarin.eu
bg.clarin.eu
cz.clarin.eu
de.clarin.eu
dk.clarin.eu
dlu.clarin.eu
ee.clarin.eu
lt.clarin.eu
nl.clarin.eu
no.clarin.eu
pl.clarin.eu
se.clarin.eu

CLARIN Portal
Offers a quick introduction to CLARIN, giving an impression about what's currently available.
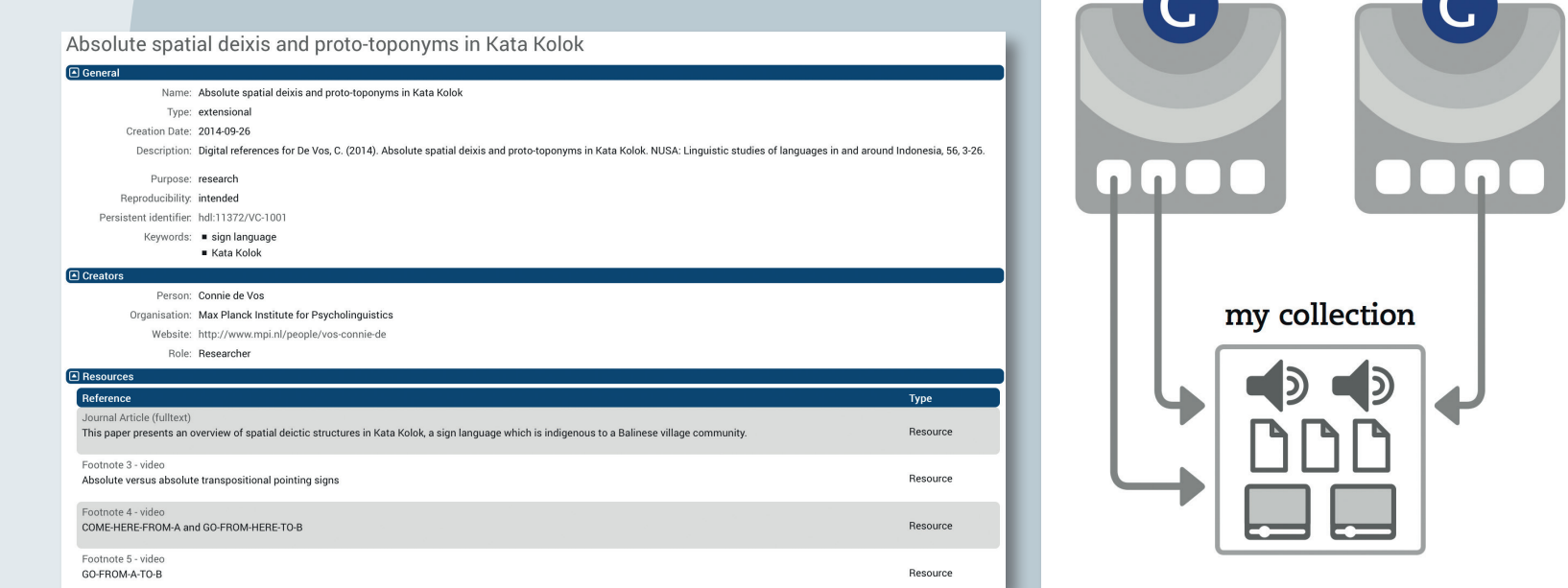
Virtual Language Observatory
Discover more than 670.000 resources and 500 tools & services using a faceted browser or a map.

Virtual Collections
Create your own digital bookmarks, ideal for citing data sets, even across repositories.
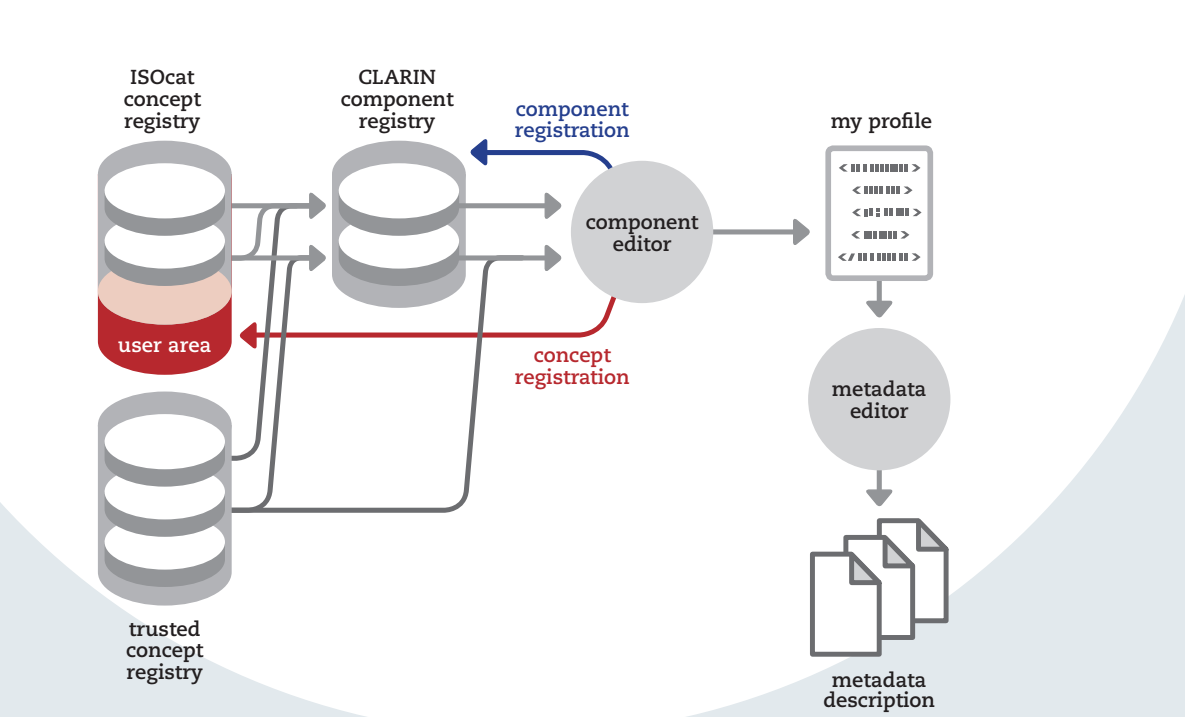
Persistent Identifiers
Persistent identifiers play a key role in keeping unambiguous references stable. CLARIN has helped to set up the service of the European Persistent Identifier Consortium.

Easy Federated Access
More than 2 million researchers and students get easy access to protected resources, with their own institutional username and password.
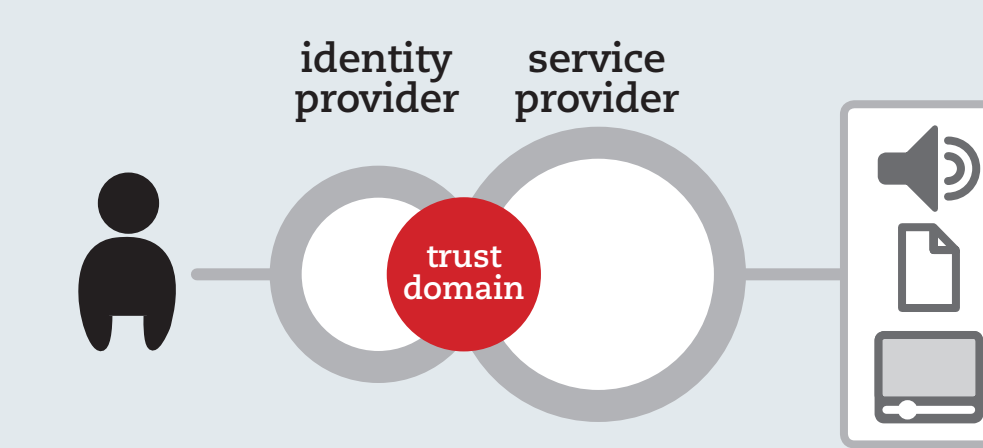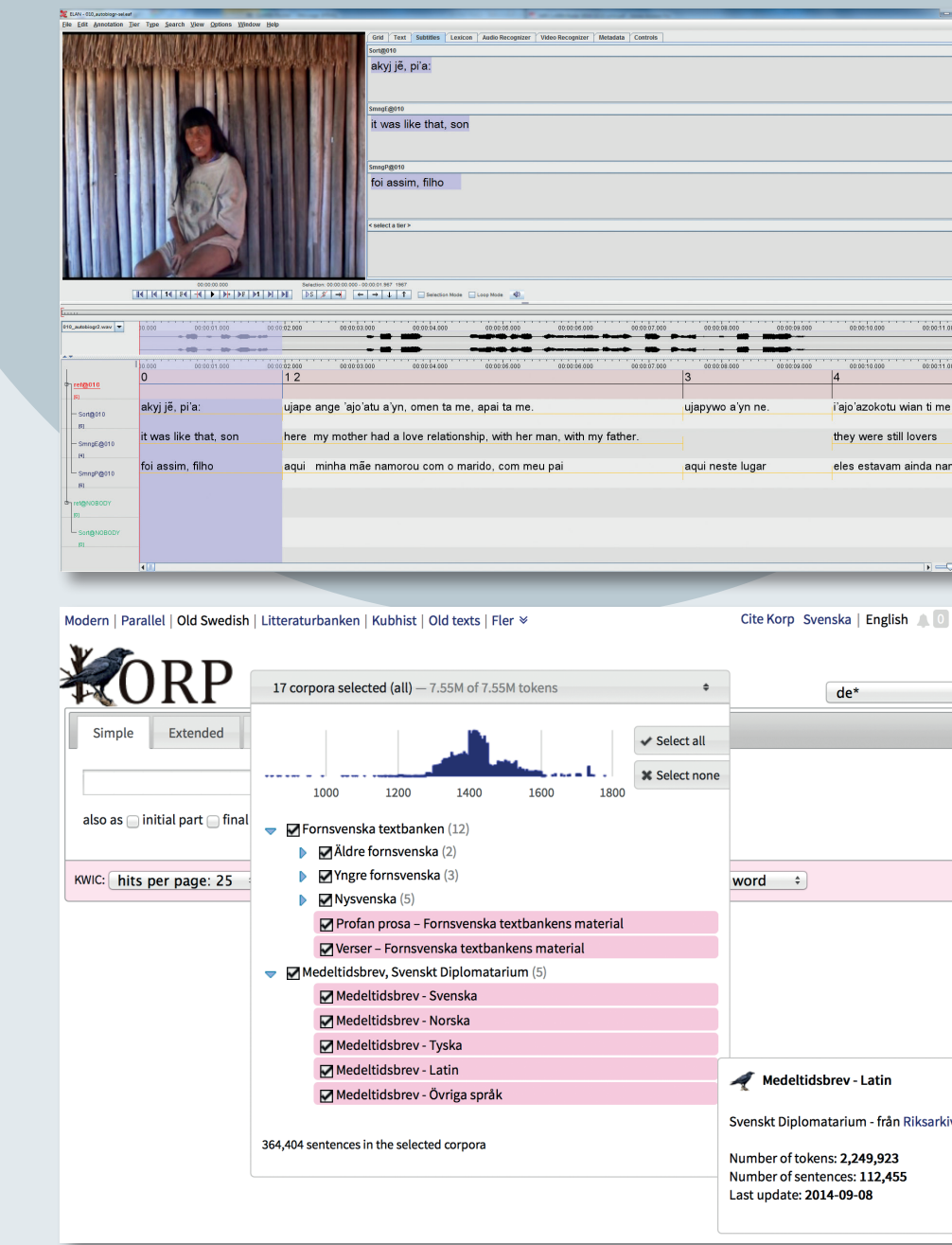
Metadata Framework
The flexible and modular component-based metadata framework CMDI allows users to describe resources and tools using domain specific terminology.
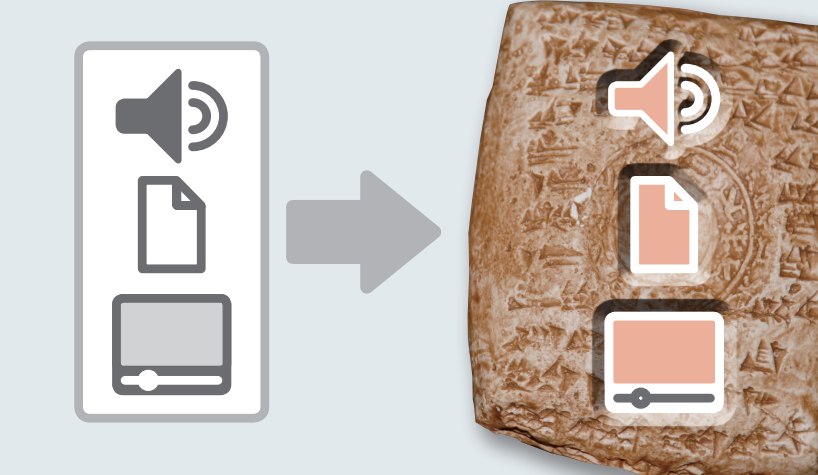
Collections and Tools
CLARIN maintains a rich set of collections (national corpora, field recordings, etc.) and widely used tools (ELAN, Exmeralda, NLP Frameworks, audio-video recognisers, WEBMaus etc.).
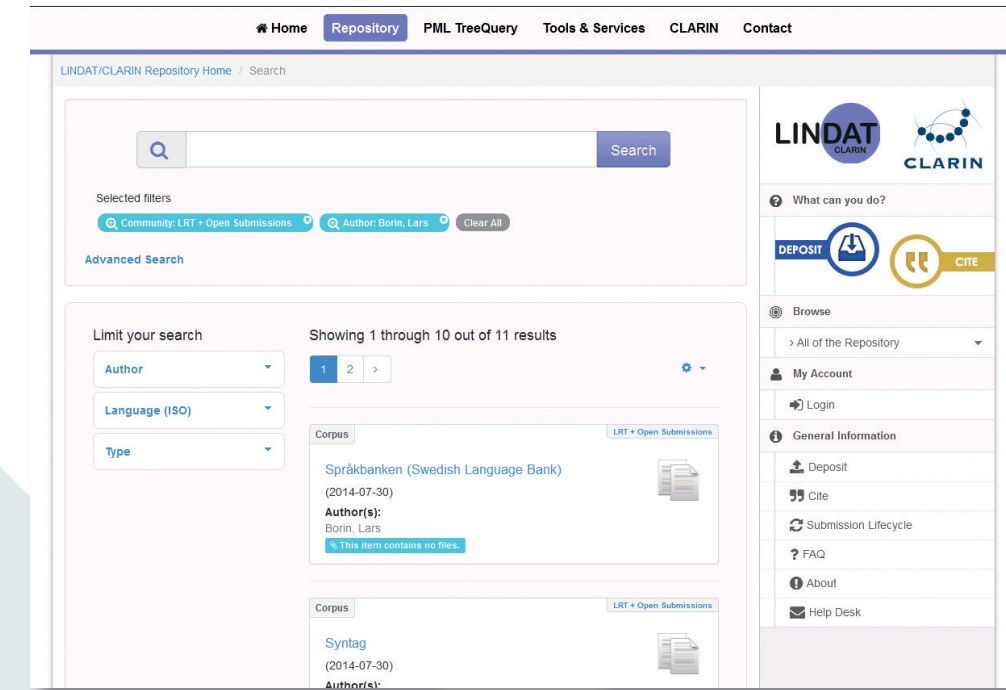
Depositing Services
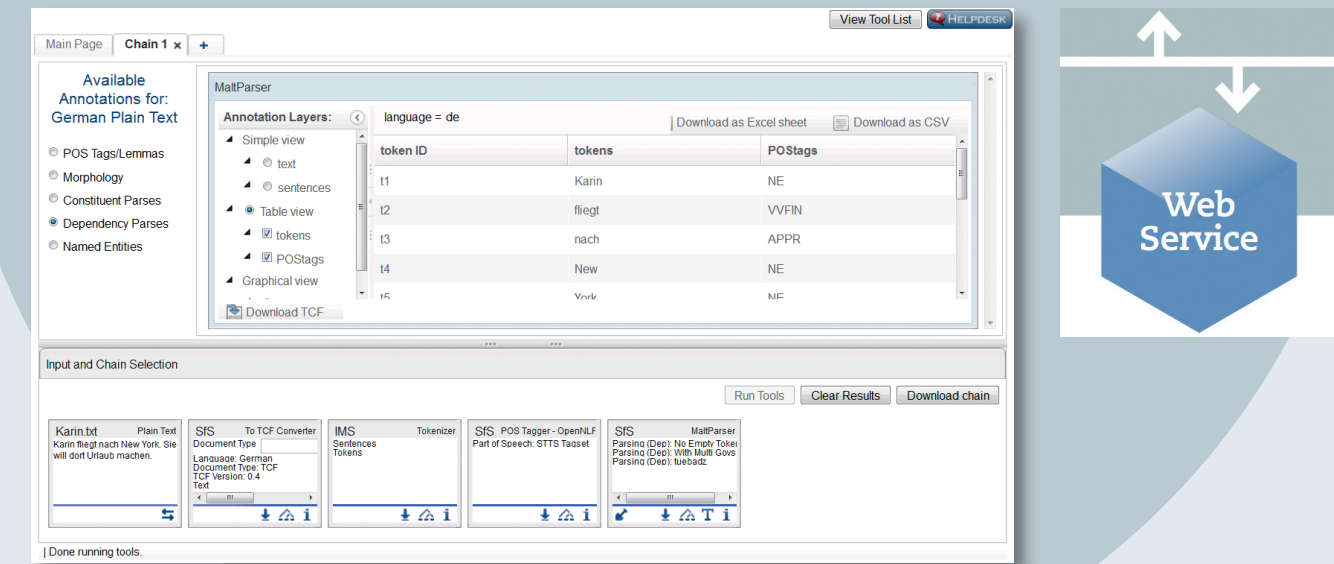Many CLARIN centres offer long-term preservation and access services.

Language Resource Inventory
Submit and access information about further language resources relevant to your research.
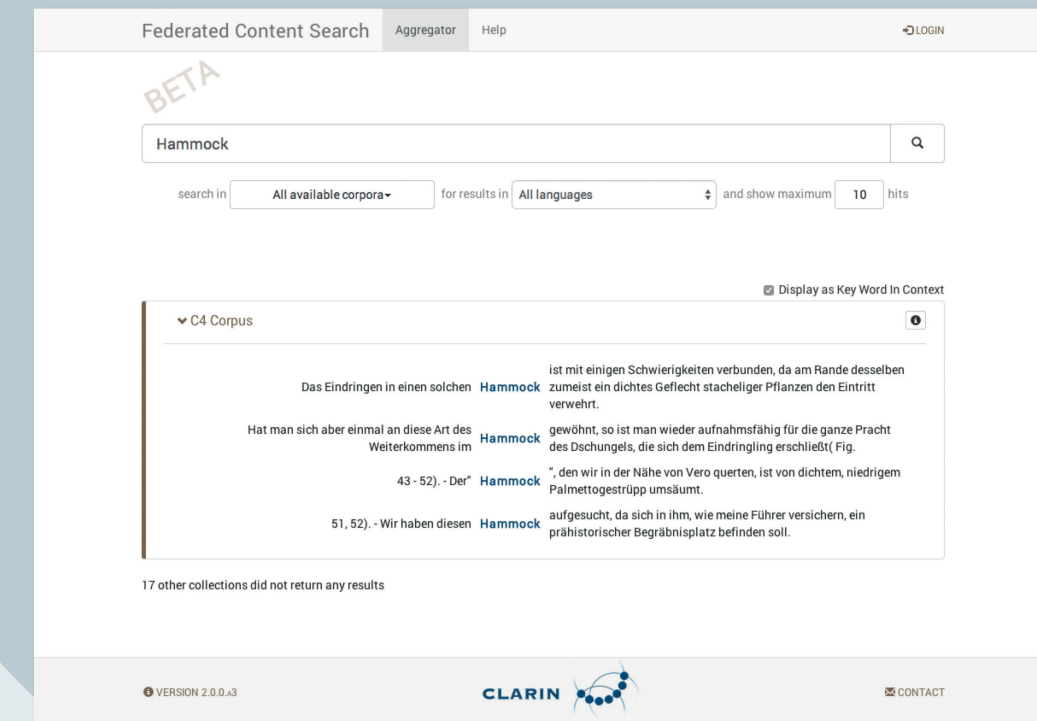
Web Services and Applications
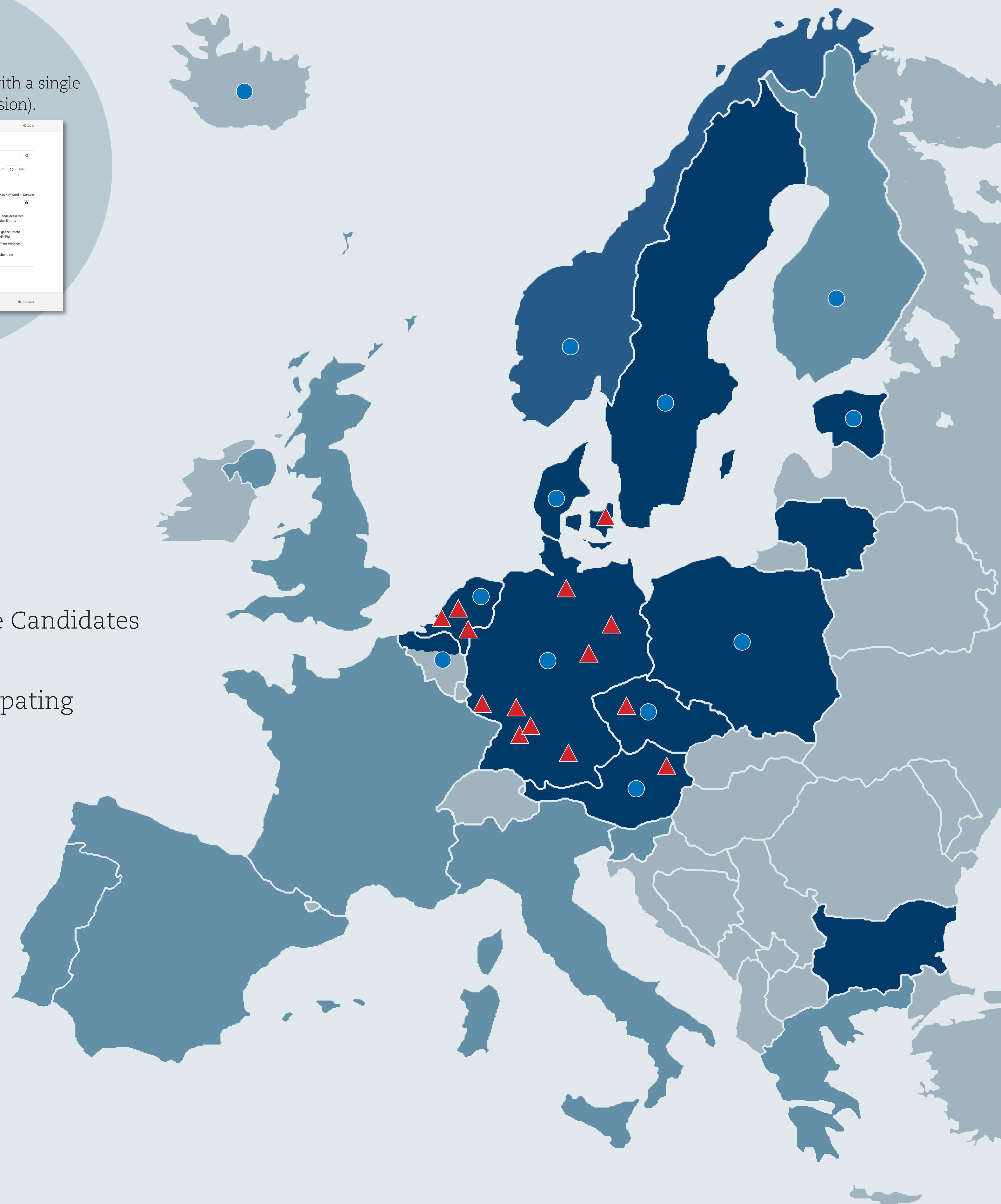Explore and analyse language data with a wide variety of tools that can be used and combined online.

Pan-European Restructuring Efforts
A growing number of repositories are currently being restructured to meet the CLARIN requirements for acting as a backbone for the infrastructure.

Content Search
Search in different corpora with a single search engine (prototype version).

- ■ Member states
- ■ Observer
- ■ Countries with Centre Candidates
- ▲ Certified Centres
- ● Countries with Participating Identity Federations

www.clarin.eu

Common Language Resources and Technology Infrastructure

A European Research Infrastructure

Language is at the heart of many disciplines in the Humanities and Social Sciences as

- an instrument for communication,
- a means to represent information and knowledge, and
- an object of study in itself.

Many collections of digital content exist, and technologies and tools have been developed to exploit, retrieve, explore, enhance and study these data. However, these resources and tools are

- geographically scattered,
- technically fragmented, and
- difficult to find, access and use.

Therefore, CLARIN is

- creating and maintaining a sustainable infrastructure
- to enable high quality, innovative digital research in the Humanities and Social Sciences.

CLARIN is based on four pillars:

- digital language content and tools,
- a technical infrastructure to provide access for the users,
- a knowledge sharing infra- structure to provide support,
- an efficient and sustainable organisational framework.

Important CLARIN principles are:

- interoperability of data and tools,
- stability and availability of services,
- long term accessibility,
- adaptability and extensibility,
- partnerships with other European infrastructures,
- embedding in the international research infrastructure land- scape, and
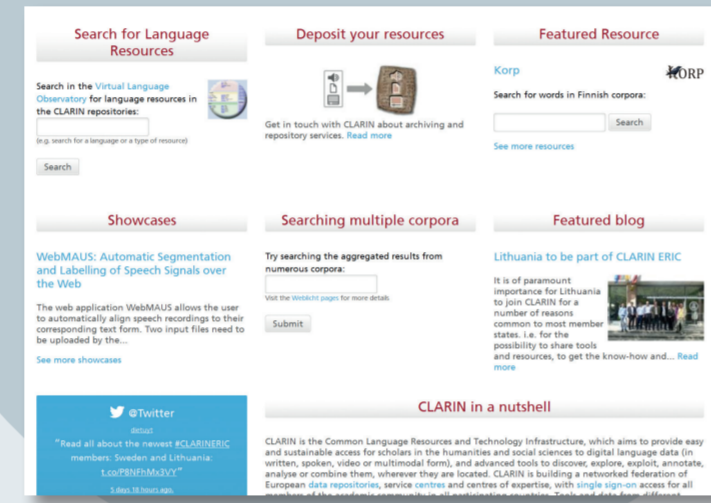- commitment to open access.

Governance and funding:

- CLARIN ERIC (European Research Infrastructure Consortium) was set up February 2012;
- founding coalition of eight countries and one intergovern- mental organisation;
- more countries have joined and are joining;
- hosted by the Netherlands;
- jointly funded by the member countries & organizations.

www.clarin.eu

at.clarin.eu

bg.clarin.eu

cz.clarin.eu

de.clarin.eu

dk.clarin.eu

dlu.clarin.eu

ee.clarin.eu

lt.clarin.eu

nl.clarin.eu

no.clarin.eu

pl.clarin.eu

se.clarin.eu

CLARIN

Towards a persistent and stable research infrastructure offering easy access to interoperable language resources and technology for the Humanities and Social Sciences

CLARIN Portal
Offers a quick introduction to CLARIN, giving an impression about what's currently available.
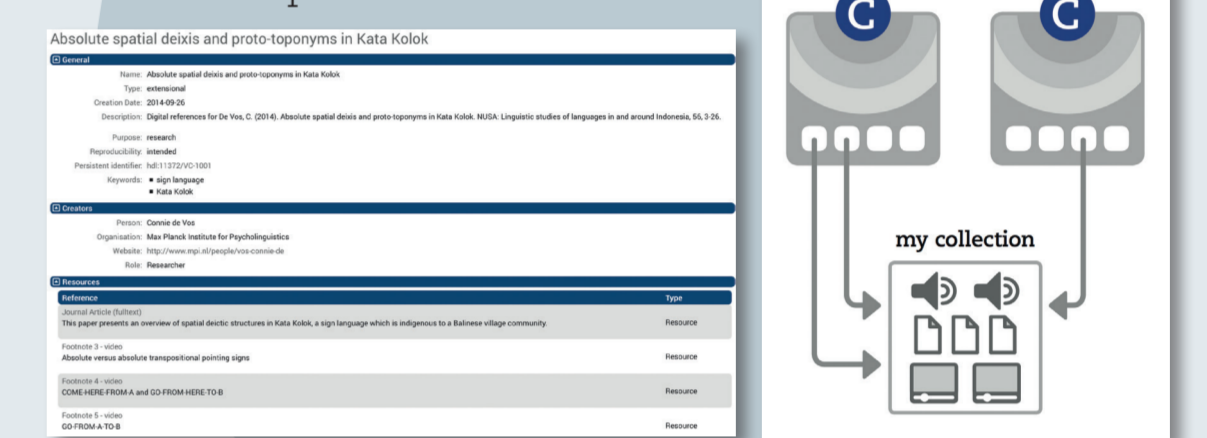
Virtual Language Observatory
Discover more than 670.000 resources and 500 tools & services using a faceted browser or a map.

Virtual Collections
Create your own digital bookmarks, ideal for citing data sets, even across repositories.
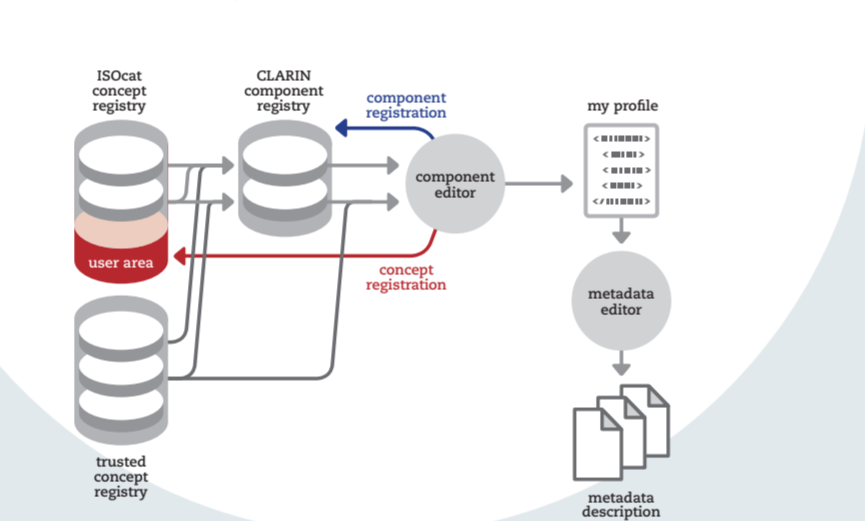
Persistent Identifiers
Persistent identifiers play a key role in keeping unambiguous references stable. CLARIN has helped to set up the service of the European Persistent Identifier Consortium.
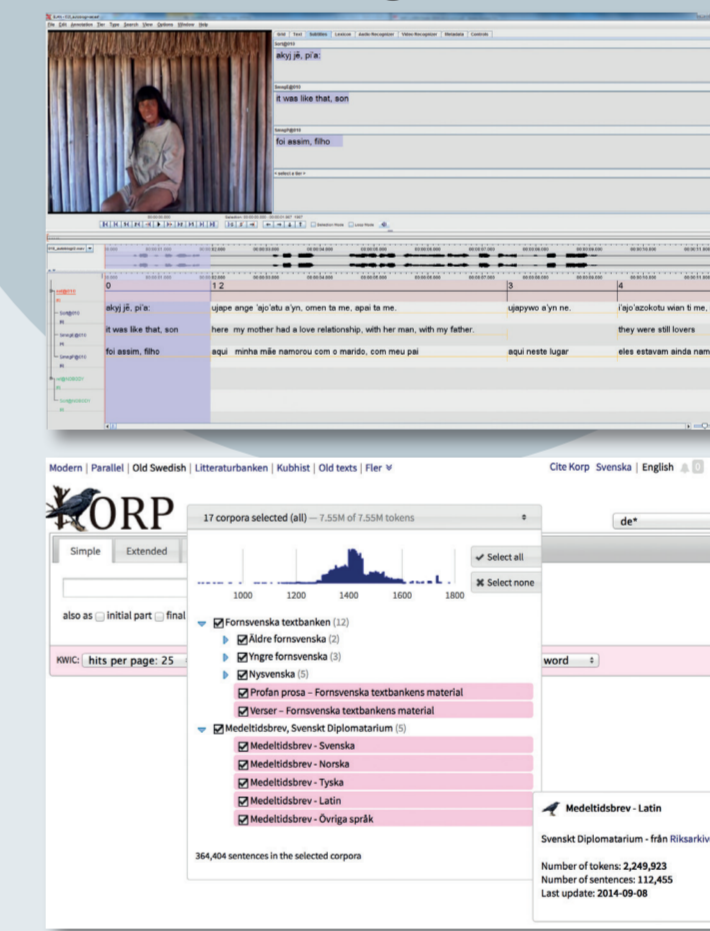
Easy Federated Access
More than 2 million researchers and students get easy access to protected resources, with their own institutional username and password.

Metadata Framework
The flexible and modular component- based metadata framework CMDI allows users to describe resources and tools using domain specific terminology.
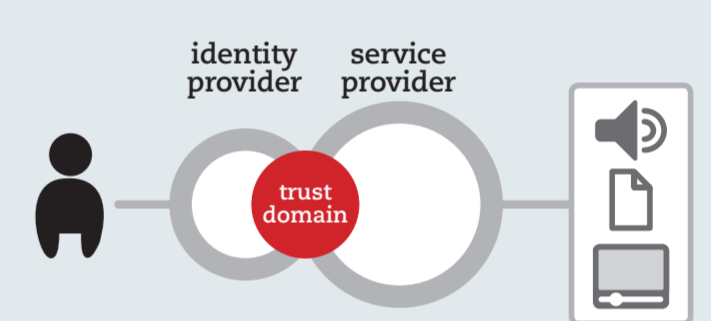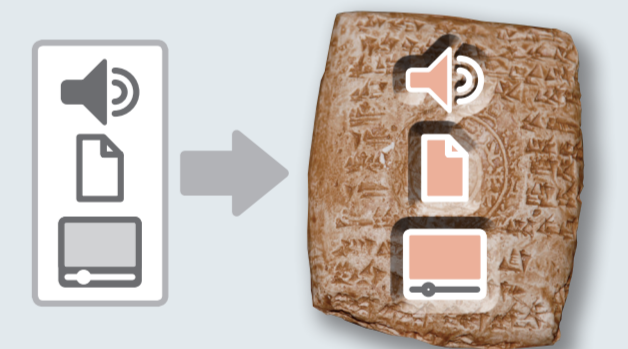
Collections and Tools
CLARIN maintains a rich set of collections (national corpora, field recordings, etc.) and widely used tools (ELAN, Exmeralda, NLP Frameworks, audio-video recognisers, WEBMaus etc.).

Depositing Services
Many CLARIN centres offer long-term preservation and access services.

Language Resource Inventory
Submit and access information about further language resources relevant to your research.
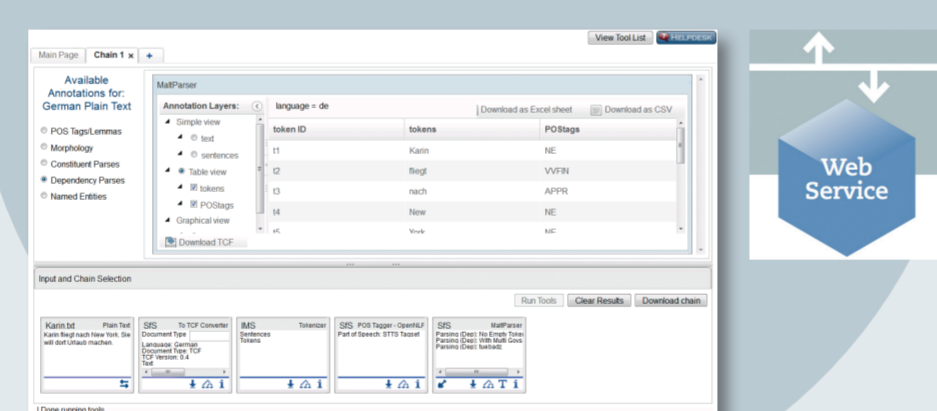
Pan-European Restructuring Efforts
A growing number of repositories are currently being restructured to meet the CLARIN requirements for acting as a backbone for the infrastructure.
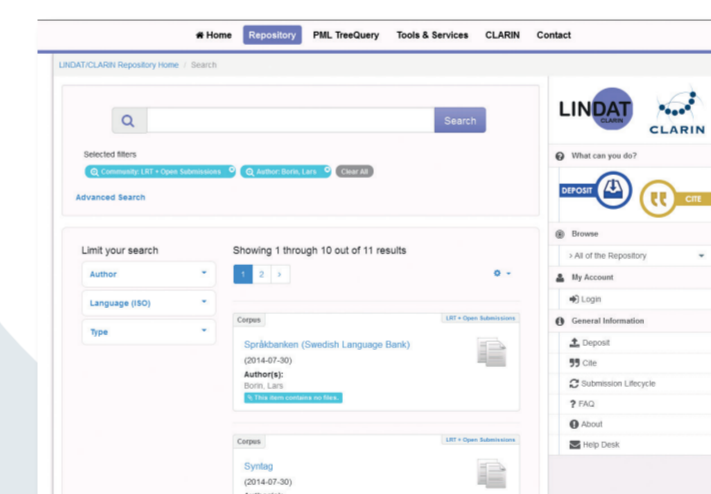
Web Services and Applications
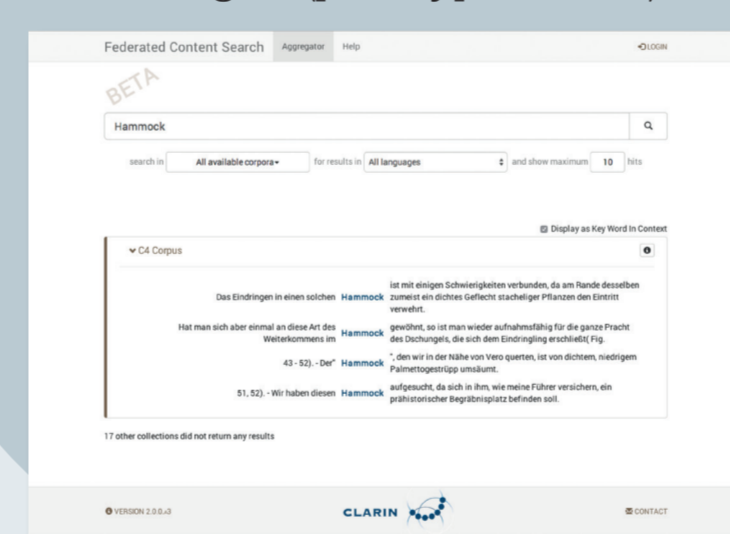Explore and analyse language data with a wide variety of tools that can be used and combined online.

Content Search
Search in different corpora with a single search engine (prototype version).

- Member states
- Observer
- Countries with Centre Candidates
- Certified Centres
- Countries with Participating Identity Federations