State of the technical infrastructure

Dieter Van Uytvanck
Technical Director CLARIN ERIC

dieter@clarin.eu

CLARIN Annual Conference 2021
Cyberspace
27 September 2021
B-centres: what happened in the last year?

• No new B-centres
• Re-assessment of 2 centres:
  – CLARIN.SI (SI)
  – ACDH-CH (AT)
• 6 more pending (waiting for CTS to be granted):
  – 2 new
  – 4 re-assessments
• in total:
  – 23 certified B-centres
  – 68 registered centres
• 29 October: deadline for the 18th assessment round
Number of centres

- B-centres
- Other centres

Year | B-centres | Other centres
--- | --- | ---
2015 | 30 | 20
2016 | 30 | 20
2017 | 30 | 20
2018 | 30 | 20
2019 | 30 | 20
2020 | 30 | 10
2021 | 30 | 10
Federated login – 2140 organisations connected
Federated login – 3163 monthly logins
New in the Virtual Language Observatory
Underneath the VLO

- The VLO has been migrated to a significantly faster server, sharing access to metadata files with
  - The harvester
  - The harvest viewer
  - The metadata curation dashboard
Metadata Curation Dashboard

• The link checker and the Curation Dashboard have been migrated to a central CLARIN ERIC server, allowing for a closer day-to-day monitoring.
• Large efforts to restructure the database and application logic.
• 5 million links checked
Virtual Collection Registry

• Four versions were released, including
  – a smoother user interface
  – new/improved functionality:
    • merging of collections
    • better programmatic access, such as OpenAPI documentation and token-based authentication

• In the pipeline:
  – copying/forking and version control
  – collaborative editing
Beware of the DOG!

Meet the **Digital Object Gateway (DOG)**:

- Framework to access Digital Objects and Collections in a uniform way
- Available as:
  - Python library
  - REST API
  - Web Application
- Compatible with CLARIN repositories, other repositories, Virtual Collections, Europeana API
- Prerelease version of the library: [https://github.com/clarin-eric/DOGlib](https://github.com/clarin-eric/DOGlib)
Language Resource Switchboard

• Version 2.3.0 of the switchboard brings improvements to the user interface, as well as several changes to the file format recognition.

• Beta version with support to extract text from HTML/PDF files

• All connected tools are now visible in the SSH Open Marketplace.
Back-end and service mobility

- Support for multiple architecture builds in our Docker compose pipeline

Source: https://www.datacenterknowledge.com/arm/its-happening-arm-server-cpus-are-coming-data-center
Yearly uptime

- 2017: 99.92%
- 2018: 99.90%
- 2019: 99.88%
- 2020: 99.92%
- 2021: 99.96%
Select your home organisation below. This is usually the organisation where you work or study. Signing in here will allow you to access certain CLARIN resources and services which are only available to users who have logged in. If you cannot find your organisation in the list below, please select the clarin.eu website account and use your CLARIN website credentials. If you don’t have such credentials you can register an account here. For questions please contact spf@clarin.eu.

Previously chosen home organisation

Utrecht University
Netherlands

Home organisation list

Search for your home organisation...

clarin.eu website account
European Union

AA@EduHr Single Sign-On Service
Croatia

Aalborg University
Denmark
Select your home organisation below. This is usually the organisation where you work or study. Signing in here will allow you to access certain CLARIN resources and services which are only available to users who have logged in. If you cannot find your organisation in the list below, please select the clarin.eu website account and use your CLARIN website credentials. If you don't have such credentials you can register an account here. For questions please contact spf@clarin.eu.

<table>
<thead>
<tr>
<th>Previously chosen home organisation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Utrecht University</td>
</tr>
<tr>
<td>Netherlands</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Home organisation list</th>
</tr>
</thead>
<tbody>
<tr>
<td>Search for your home organisation...</td>
</tr>
<tr>
<td>All countries</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>clarin.eu website account</th>
</tr>
</thead>
<tbody>
<tr>
<td>European Union</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>AAI@EduHr Single Sign-On Service</th>
</tr>
</thead>
<tbody>
<tr>
<td>Croatia</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Aalborg University</th>
</tr>
</thead>
<tbody>
<tr>
<td>Denmark</td>
</tr>
</tbody>
</table>
Revamped website

Knowledge Infrastructure

The mission of the CLARIN Knowledge Infrastructure is to ensure that the available knowledge and expertise are made accessible.

Ready-to-use Language Resources

CLARIN enables research based on digital language resources by offering advanced services to discover, explore, exploit, annotate, analyse, combine or archive language data.

Read more
Revamped website

CLARINO Bergen Center

The CLARINO Bergen Center provides access to language data and analysis tools through a repository, a corpus management and search platform, a treebanking platform, and a component metadata editor. It is also integrating access to a Term Portal, a Medieval Nordic Text Archive, and other collections.

Featured example: COMEDI

Visit tool portal

PORTULAN CLARIN

The PORTULAN CLARIN workbench makes available a wide range of language processing tools, such as POS tagging, named entity recognition, sentiment analysis, dependency parsing, among others. These tools are made available through different modes of interaction, namely through a browser interface, a file processing interface and a web service API.

Featured example: LX-DepParser

Visit the workbench
Revamped website – please share your suggestions

CLARIN – the research infrastructure for language as social and cultural data

CLARIN is a digital infrastructure offering data, tools and services to support research based on language resources.
Thank you for your attention!

With special thanks to:

• The assessment committee:
  – Lene, Cyprian, Daan, Jozef, Riccardo, Tomas
• The CLARIN ERIC central developer team:
  – Alex, André, Menzo, Michał, Twan, Willem
• Developers and contributors from the national consortia
• The taskforces:
  – AAI, CMDI & Curation, FCS, PIDs
• And all the others who are contributing to the construction & operation of the infrastructure!