



State of the technical infrastructure

Dieter Van Uytvanck
Technical Director CLARIN ERIC
dieter@clarin.eu

CLARIN Annual Conference 2022
Prague & Cyberspace
11 October 2022

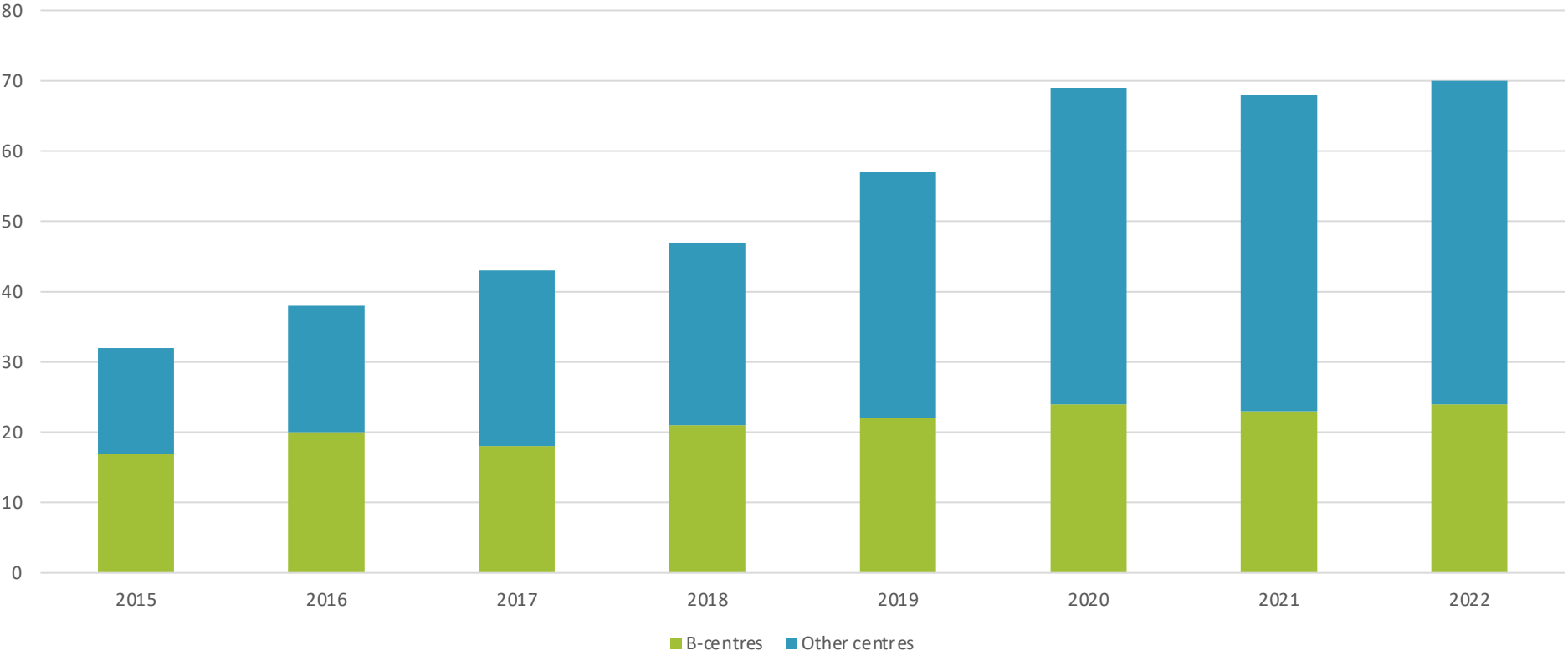
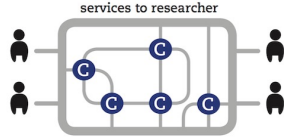
Back to ~~normal~~ better!

- Welcome back after two virtual CLARIN conferences
- Let's embrace the positive aftermath of the pandemic episode: a fully hybrid annual conference

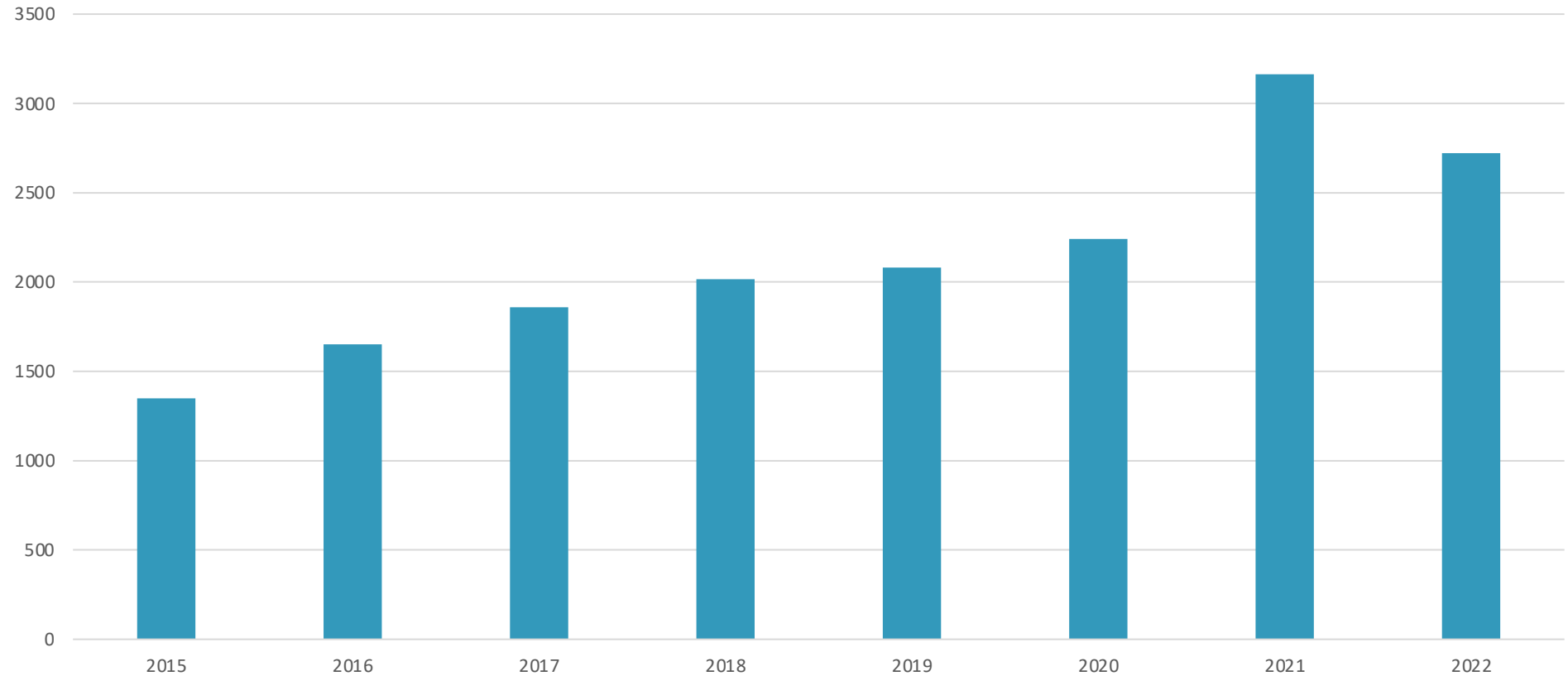
B-centres: what happened in the last year?

- New B-centre: CLARIN:EL (GR)
- Re-assessment of 3 centres:
 - CMU-Talkbank (US)
 - The Language Bank of Finland (FI)
 - MPI for Psycholinguistics (NL)
- more pending (waiting for CTS to be granted)
- 31 October: deadline for the **20th** assessment round

24 B-centres, 70 in total



2722 monthly federated logins



Virtual Language Observatory



VLO / Faceted search / Record: The Zurich Tangram Corpus - UZH Edition

Record 6 of 546,390

< previous next >

The Zurich Tangram Corpus - UZH Edition

Record details Links (88) Availability All metadata Technical Details Hierarchy

Name	The Zurich Tangram Corpus - UZH Edition
Description	This corpus contains tasks, where one subject (the instructor) describes different Tangram figures to another subject (the receiver) so that the receiver can recreate the same order of figures that the instructor has in front of them. The subjects initially don't know each other and work together to solve these tasks in three consecutive sessions. This edition features the complete recordings, but lacking phone and word segmentation. Subjects audio tracks are combined into stereo files. If you would like just the transcribed segments with separate files for the subjects or want the word and phone segmentation see corpus ZTC_BAS.
Collection	Bavarian Archive for Speech Signals (BAS)
Language	English German
Modality	spoken
Continent	Europe
Country	Switzerland
Genre	tangram figures
Subject	tangram figures
Organisation	Bavarian Archive for Speech Signals, Ludwig-Maximilians-Universität München
National project	CLARIN-D
Resource type	audio
Data provider	Bayerisches Archiv für Sprachsignale

Record identifier HDL 11022/1009-0000-0007-D838-7@format=cmdi



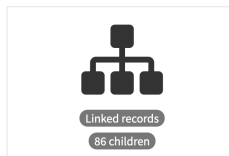
HDL 11022/1009-0000-0007-D...

Landing page

Linked resource

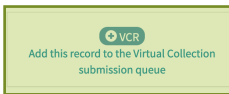


CLARINDocu.zip



Linked records

86 children



+ VCR
Add this record to the Virtual Collection submission queue

Hide

Items to submit to the VCR

- Iwo Gall -teksty teatralne
- PhonDat 1

Submit

Clear

Metadata Curation & Link Checker



- A lot of work on the [link checker](#) and the [Curation Dashboard](#) took place
- **Restricted access** now a separate category
- Link checker now also available for testing via an (alpha) API
- Visit Wolfgang's bazaar stall this afternoon to learn more...

Overall

Category	Count	Average Response Duration(ms)	Max Response Duration(ms)
Ok	1081684	923.8	5566499.0
Undetermined	549	15.1	36.0
Restricted_Access	565818	235.8	4776.0
Blocked_By_Robots_txt	32694	N/A	N/A
Broken	50231	2235.8	6643.0
Invalid_URL	1	N/A	N/A
Total Count: 1730977 · Average Response Duration(ms): 752.8 · Max Response Duration(ms): 5566499.0			

Virtual Collection Registry



- Seven versions were released, including
 - a smoother user interface
 - improved connectivity
 - CMDI explorer (collection download)
 - Switchboard pop-up
 - Easy javascript-based widget for integration

VCR integration example

Resources

Some resource

<http://doi.org/10.5555/12345678>

Add to virtual collection

Other resource

<https://tinyurl.com/my-resource>

Add to virtual collection

One more resource

<https://github.com/clarin-eric/js-vcr-integration/>

Add to virtual collection

Language Resource Switchboard



- Versions 2.4.0 and 2.4.1 of the Switchboard bring
 - improvements to the user interface
 - (experimental) support for multiple input files

Switchboard ⌵ ✕

3023.txt 463.27 KiB Show content

Mediatype text/plain ▾

Language English ▾

Matching Tools

Group by task

▾ Constituency Parsing

> WebLicht Const Parsing EN Open Requires authentication

▾ Dependency Parsing

> UDPipe Open

Bringing it all together (1)



- Which virtual collections contain at least one broken link?
- Team VCR & VLO & link checker [can help out!](#)

Virtual Language Observatory Search Contributors Help CLARIN

VLO / Faceted search / Search results

Showing 2 results within selection for Results per page: 10

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

Collection CLARIN Virtual Collection Registry [more...](#)

Format

Keyword

Organisation

CLARIN services in the European Open Science Cloud

(Part of CLARIN Virtual Collection Registry)

Links to data sets and services that are connected via the Language Resource Switchboard.; Contains a link to a digitized pamphlet as PDF, with embedded text. This file is processed with Voyant.; Contains a link to the plain text version of the pamphlet. This file is processed with WebLicht and UDPipe.; Overview of CLA...

[Landing page for this record](#)

Tutorials for the Switchboard tools

(Part of CLARIN Virtual Collection Registry)

This collection contains existing tutorials and guides for the tools and services listed in the Switchboard Tool Inventory. The collection is useful for researchers, teachers and students who are interested in using the Switchboard tools in research, teaching and learning.

[Landing page for this record](#)

Bringing it all together (2)

Tutorials for the Switchboard tools



Record details

Links (32)

Availability

All metadata

Technical Details

Name	Type
 introducing-udpipe-for-easy-natural-language-processing-in-r	HTML ⋮ ➤
 udpipe-annotation.html	HTML ⋮ ➤
 userguide.html	HTML ⋮ ➤
 intro.html	HTML ⋮ ➤
  voyant-tools-a-tutorial-for-text-analysis-df265d85d214	HTML ⋮ ⌵

Resource details

Mime type: text/html

Link: <https://medium.com/@priestleyal/voyant-tools-a-tutorial-for-text-analysis-df265d85d214>

Automatic link checking

 **The resource may not be available at this location.**

Result of automatic checking: the resource is unavailable (410 Gone)

Last check: 7 October 2022, 09:01

CLARIN for programmers - clarin.eu/programmers

- APIs
 - Central OpenAPI documentation for VCR, DOG, linkchecker
 - More future work on API-based design, e.g. VLO 5.0
- Javascript widgets for easy integration
 - Language Resource Switchboard
 - Virtual Collection Registry
- Notebooks
 - landing page: clarin.eu/notebooks
 - See *Processing Europeana Text Collections with Jupyter Notebooks* by Twan Goosen and Michał Gawor – this afternoon, *Teaching with CLARIN* session

Renewed status.clarin.eu

CLARIN service status



Subscribe to Updates

✓ All Systems are Online

[Website](#) | [Support](#) | [Uptime statistics](#)

Overview

- CLARIN website Good Service
- Centre Registry Good Service
- Component Registry Good Service
- Discovery Service Good Service
- Federated Content Search Good Service
- Language Resource Inventory Good Service
- Language Resource Switchboard Good Service
- CLARIN Identity Provider (user.clarin.eu) Good Service
- Virtual Collection Registry Good Service
- Virtual Language Observatory Good Service

Upcoming Events

Next 30 Days

There are no upcoming events.

[» View More](#)

Recent Events

Last 7 Days

Thursday, October 6th 2022

There are no reported events.

Wednesday, October 5th 2022

VLO maintenance

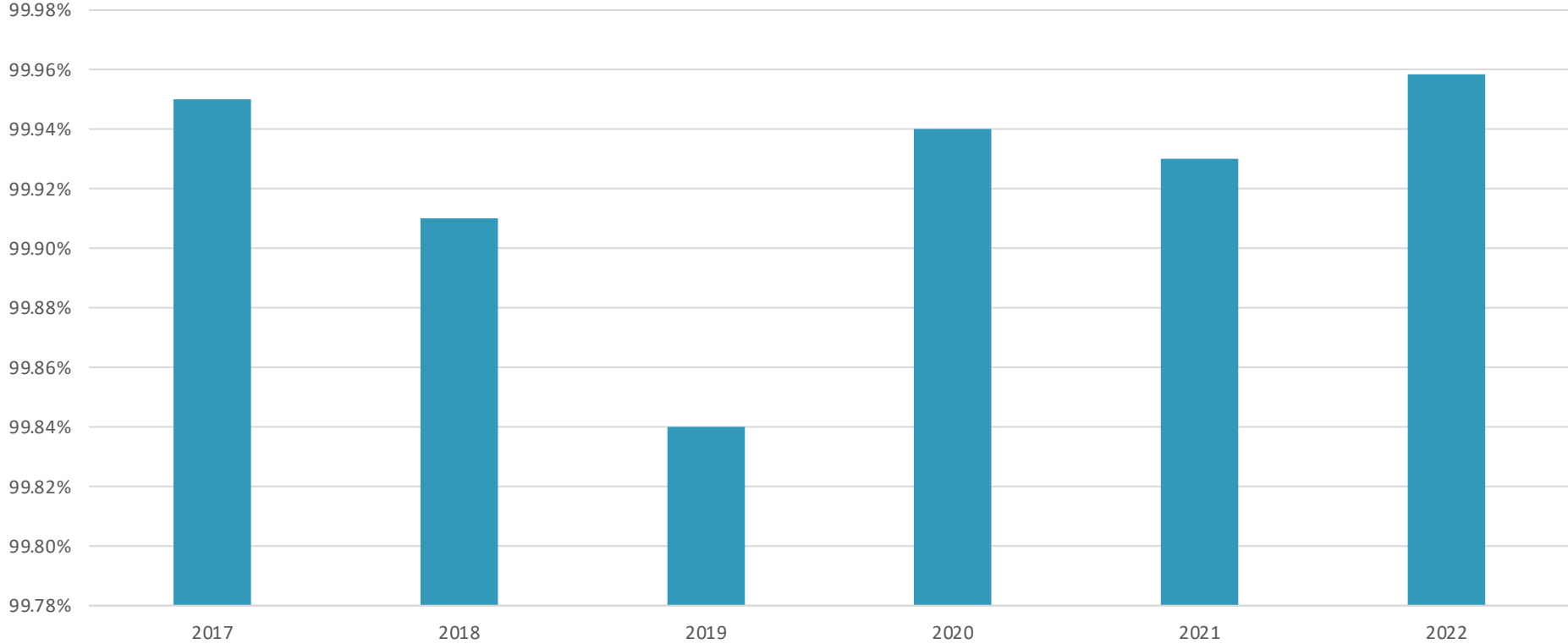
[» View Event Details](#) | Created Wednesday, 5th October 2022 11:00 CEST

Scheduled

We will be performing maintenance to the Virtual Language Observatory on Wednesday, 5 October 2022, between 11:00 and 12:00 CEST. As a result, the VLO (at <https://vlo.clarin.eu>;) will not be available and/or may show incomplete content for some time within this slot.

Posted: Wednesday, 5th October 2022 11:00 CEST

Yearly uptime: 99.96%



Looking into the future



Looking into the future

- The illustration on the side is DALL-E's artistic interpretation of human language:
 - *Art deco Prague illustration with Charles bridge fading into a lit sci-fi cyber highway*
- Well suited since our host LINDAT/CLARIAH-CZ plays a major role in the development of language AI data sets and tools



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, Patrick, Twan, Willem, Wolfgang
- Developers and contributors from the national consortia
- The taskforces:
 - CMDI & Curation, FCS, PIDs
- And all the others who are contributing to the construction & operation of the infrastructure!

One more thing...

Franciska's special CLARIN resource family

Limited Edition ... Unlimited Fun!



CLARIN 
RESOURCE FAMILIES