

State of the technical infrastructure

Dieter Van Uytvanck

Technical Director CLARIN ERIC

dieter@clarin.eu

CLARIN Annual Conference 2016

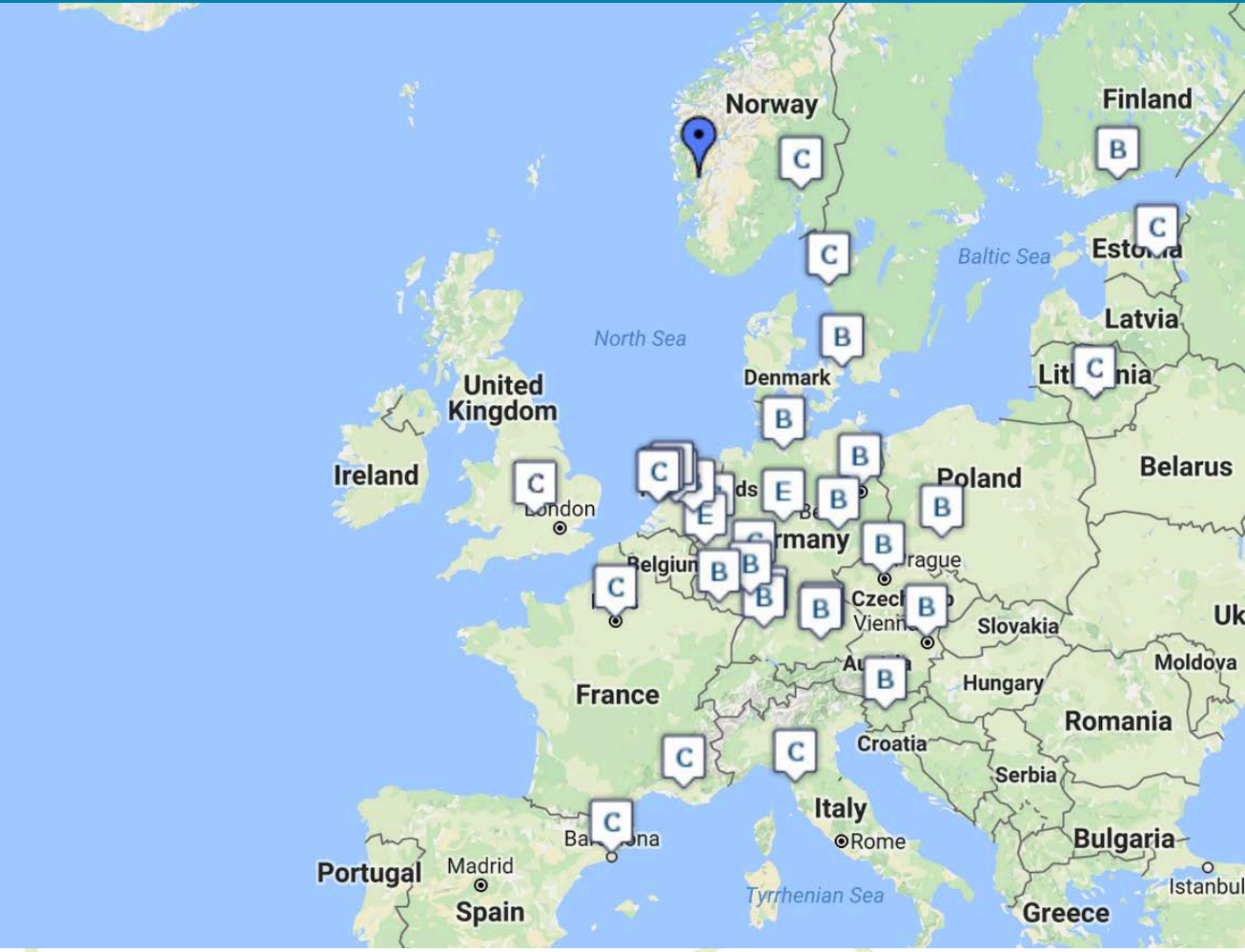
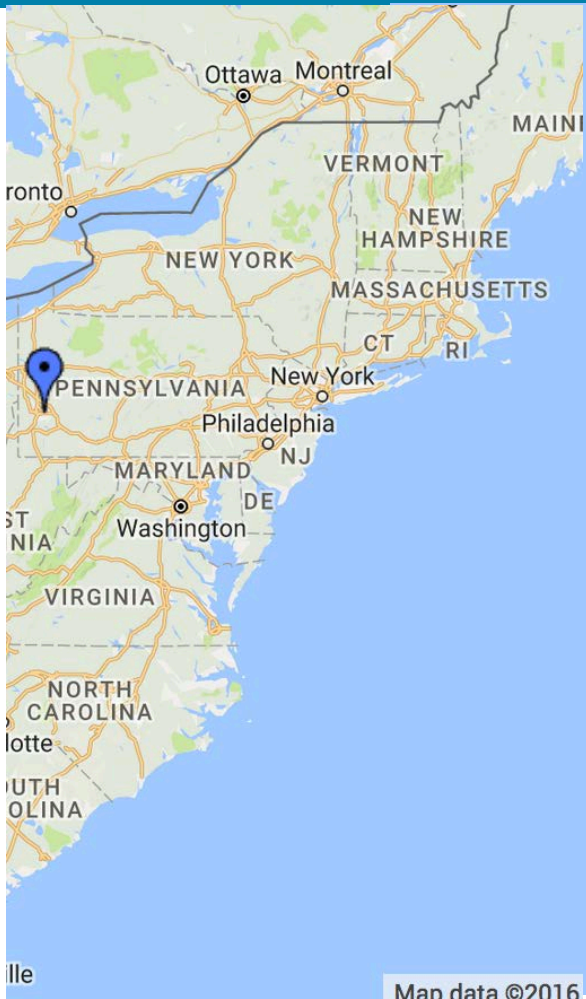
Aix-en-Provence

28 October 2016



Centres

- 3 new B-centres : Language Bank (FI), CLARIN.SI (SI), CLARINO Bergen (NO)
- successful re-assessment of LINDAT (CZ)
- several re-assessments postponed due to new Data Seal of Approval guidelines
- in total:
 - 20 certified B-centres
 - 38 registered centres



A-services

- LINDAT's LRT inventory is the first A-service

The screenshot displays the LINDAT LRT inventory interface. At the top, there is a search bar with a magnifying glass icon and a 'Search' button. Below the search bar, the 'Selected filters' section shows 'Community LRT + Open Submissions' with a 'Clear All' button. The 'Advanced Search' section is also visible. On the right side, the LINDAT and CLARIN logos are present, along with a 'What can you do?' section containing 'DEPOSIT' and 'CITE' buttons. Below this is a 'Browse' section with a dropdown menu set to 'All of the Repository'. The 'My Account' section includes a 'Login' button. The 'General Information' section lists 'Deposit', 'Cite', 'Submission Lifecycle', 'FAQ', 'About', and 'Help Desk'. The main content area shows 'Limit your search' filters for 'Author' and 'Subject'. The 'Subject' filter is expanded, showing categories like 'Germanistik (47)', 'corpus (8)', 'written language (2)', 'balanced corpus (3)', 'APICES (2)', 'classroom (2)', 'informal spoken language (2)', 'language acquisition corpus (2)', and 'youth language (2)'. The search results section indicates 'Showing 1 through 10 out of 988 results' and includes a pagination control. The first result is 'jps100k 2.0' by 'Erzenc, Tomaz; Filip, Darja; Koz, Simon; Ledinek, Nina', with a 'Publicly Available' status and a 'This item contains 1 file (4.28 MB)' note.

Federated login, 2016

Connected / Planned / Possible

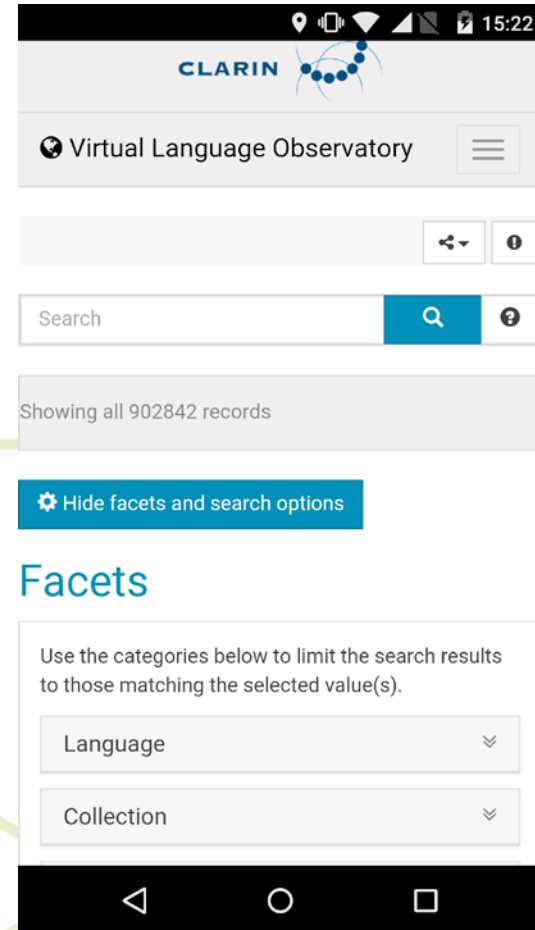
1. ACOnet, Austria
2. Belnet Federation, Belgium
3. SWITCHaai, Switzerland
4. AAI@EduHr, Croatia
5. eduID.cz, Czech Republic
6. WAYF, Denmark
7. Deutsches Forschungsnetz (DFN), Germany
8. **TAAT, Estonia**
9. SIR, Spain
10. Haka, Finland
11. Fédération Éducation-Recherche, France
12. GRNET, Greece
13. eduID.hu, Hungary
14. Edugate, Ireland
15. **IDEM, Italy**
16. **LAIFE, Latvia**
17. SurfConext, The Netherlands
18. FEIDE, Norway
19. **PIONIER.id, Poland**
20. RCTSaai, Portugal
21. SWAMID, Sweden
22. ArnesAAI, Slovenia
23. UK Federation, United Kingdom
24. InCommon, United States of America
25. LITNET, Lithuania

Federated login

- New countries in Service Provider Federation: Estonia, Italy, Latvia, Poland
- Total number of organisations that can login: 1230+
- On average 55 logins per day via the central discovery service
- Redundant setup for Discovery Service

Virtual Language Observatory

- New versions with improved functionality
 - hierarchies in 3.3
 - availability and license information in 3.4
 - complete (responsive) design overhaul in 4.0
- Efforts to curate the metadata via CLARIN-PLUS



(Federated) Content Search

- New version of the protocol, to support CQP-alike queries, is finalized
- New prototype version of the content search engine released
- The first 2 endpoints supporting the new protocol are up

The screenshot displays the Multi-layer FCS query interface. At the top, a search bar contains the query `[word = 'Surovy'] [pos = 'VERB']`. Below the search bar, there are dropdown menus for 'Any Language', 'Multi-layer Federated Content Search Query Language (FCS-QL)', and 'All available collections'. A control for 'and show up to 10 hits per endpoint' is also visible. The results section shows '1 matching collections found' and a checkbox for 'Display as Key Word In Context'. The first result is from 'The Språkbanken corpora - Unknown Institution' and contains several sentences with the word 'Surovy' highlighted in blue. The sentences are: 'Ja, Tomas Surovy drog till NHL .', 'Sudden vidtar och där tar det inte lång tid innan Tomas Surovy vråker in 6-5 med en avvaktande Morautvisning .', 'Hoppas dock att man satt in nya burar sedan han lärde sig var alla maskorna satt förra säsongen ... Även om fokus ligger mycket på Surovy får vi hoppas att Luleå lyckas fokusera på annat än skitiga domslut och fult spel som inte beivras , för det blir det .', 'På isen får dom ju gärna visa var hjärtat hör hemma ... Jag tror att dom gör det och vinner över Surovy Lions ... Undrar hur Surovy tänker ... ?', 'Surovy hade den glädjen att bli pappa här i natt , tror jag det var .', and 'Sorry Therese ... HV har egen is att lira på och vilken temp den än har så finns det en trygghet i det och även runt om och som det såg ut senast på det andra lagets is ... Hade man bara startat samtidigt vet då inte jag hur det hade gått ... Ikväll vinner HV med 5-3 ... Jag fick en skön kommentar av " Osten " , han har aldrig uppgett vilken sort men den är garanterat Ädel ... Han skrev ... ; - " För varje omgång i KS ökar insikten om att Surovy gjorde sin viktigaste insats för Luleå iklädd LHC tröja .'

Language Resource Switchboard

- The missing link between metadata & data (files) & tools (web applications)

VLO / Faceted search / Search results / Record: story_biruma



Record 2 of 117

< previous

next >

story_biruma

Show the original provider's page for this record



Record details

Resources (4)

Availability

All metadata

Technical details

Hierarchy

Name	Type	
 story_biruma.wav	audio file	... >
 story_biruma.mpg	video file	 Process with Language Resource Switchboard >
 story_biruma.mpeg	video file	... >
 story_biruma.tbt	annotation file	... >

Back-end services, service mobility, migration (1)

- the CLARIN Concept Registry and CLAVAS are being migrated to OpenSKOS 2.0 (CLARIN-PLUS)
- CMDI 1.2
 - was released and is now
 - integrated into the Component Registry
- Service mobility:
 - Almost all central CLARIN services are now available as a (Docker) container
 - much work has been done to setup a standardized deployment workflow
- Migration of the CLARIN Identity Provider is on the way (CLARIN-PLUS)

Back-end services, service mobility, migration (2)

- A-service requirements (including easy migration to other centres) have been settled
- Human Interface Guidelines are used for the CLARIN website and the VLO, other applications will follow gradually (CLARIN-PLUS)
- Report on risk management for eInfrastructures has been written (CLARIN-PLUS)

status.clarin.eu

🔄 Announcements

This is an overview of central CLARIN services and A-services. Please see www.clarin.eu/status for details on service issues and planned maintenance.

📈 7 Day Uptime

99.99%

🔄 Uptime History

Status	Name	Rate	Oct 21	Oct 22	Oct 23	Oct 24	Oct 25	Oct 26	Oct 27
🟢	CLARIN website	1m	100%	100%	100%	100%	100%	100%	100%
🟢	Centre Registry	1m	100%	100%	100%	100%	100%	100%	100%
🟢	Component Registry	1m	100%	100%	100%	100%	100%	100%	100%
🟢	Discovery Service	1m	100%	100%	100%	100%	100%	100%	100%
🟢	Federated Content Search	1m	100%	100%	100%	100%	100%	100%	100%
🟢	Language Resource Inventory	1m	100%	100%	100%	100%	100%	100%	100%
🟢	user.clarin.eu	1m	100%	100%	100%	100%	100%	100%	100%
🟢	Virtual Collection Registry	1m	100%	100%	100%	100%	100%	100%	100%
🟢	Virtual Language Observatory	1m	100%	100%	100%	100%	100%	99.64%	100%

The connection to EUDAT

- A training workshop was organized in December 2015 about the use of B2SAFE (iRODS-based safe replication)
- Around 8-10 centres expressed interest
- The implementation is now taking place
- Planned for 2017: a plugin for the DSpace repository

Documentation & quality guidelines

- Software guidelines
- Testing pool (see clarin.eu)

wiki: [SoftwareDevelopment](#) / [ProjectGuidelines](#)

[Up](#) | [Start Page](#) | [Index](#) | [History](#)

Author/responsible: [Twan](#)

Guidelines for software development in projects

This documents contains guidelines for software development within CLARIN projects. Since such projects generally have a limited temporal scope, and the original developer(s) will often not be available for support outside the context of the project, additional care has to be taken with respect to a number of aspects that factor into the 'sustainability' of a software project. Providing good documentation, choosing commonly understood and well supported technologies, guarding quality throughout the development process, and ensuring the portability of the resulting code base strongly increase the chances of a piece of software surviving the end of its "parent project".

We (CLARIN) suggest that you commit to these guidelines if you are working on a software project in the context of a project. You can do this explicitly by having this document referenced in the relevant task description(s).

[Guidelines for software development in projects](#)
[Documentation](#)
[Implementation decisions](#)
[Quality and integrity](#)
[Portability](#)
[Other](#)
[Resources](#)

Acknowledgements

- The assessment committee: Lene, Daan, Patrice, Jozef, Tomas
- The CLARIN ERIC central developer team: Menzo, Sander, Twan, Willem
- Developers from the national consortia: Thomas, ...
- The taskforces: AAI, CMDI, Curation, FCS, PIDs
- The CLARIN-PLUS development team: Amir, Pavel, Davor, Go, Matej, Tomasz, Olha, Claus, Leif-Jöran, Olga, Kairi, Kadri, Oliver
- And all the others who are contributing to the construction of the infrastructure!

A prediction by Cézanne?

(1866)



A prediction by Cézanne?



le nouvel
Observateur

Thank you for your attention!

- Questions?