

CLARIN 2021

Presentations by CLARIN Committees

Chair: Franciska de Jong

Day 3

Wednesday 29 September

11:00 - 11:30



Assessment Committee CAC 2021

Lene Offersgaard (chair), DK

Jozef Misutka(vice chair), CZ

Cyprian Laskowski, SI

Daan Broeder, NL

Riccardo Del Gratta, IT

Tomas Krilavičius, LT

B-Centre Assessments Status

(period: mid-Sep 2020 – mid-Sep 2021)

No new B-centres

Re-assessment of 2 centres:

- CLARIN.SI Language Technology Centre (Sep 30, 2020)
- ACDH-CH - Austrian Centre for Digital Humanities and Cultural Heritage (July 9, 2021)

In assessment process: 3 (round 17)

Centres still in the CTS process in total: 6

- Getting the CTS easily takes up to 9 months!

Steps to Become a B Centre:

1. First become a C-centre (metadata-harvest to VLO)
2. Fill in the application for CTS (CLARIN will pay the fee)
3. Fill in the [checklist](#) - questions always welcome
4. Apply for B-centre certification: send checklist and attach your CTS application when applying
5. Be prepared to answer questions from us
6. Prepare feedback to CTS, when requested
7. Tell us when you get the CTS

Mail us: assessment@clarin.eu

Next Rounds

Next Deadlines: October 29, 2021

April 29, 2022

The B-centres with a positive assessment will receive the B-centre label when they get the Core Trust Seal.

Validity

The B-centre status is synchronised with Core Trust Seal validity: 3 years

Checklist will be updated in January 2022

Use new checklist for deadline in April 2022

We welcome feedback on the checklist

Mail us: assessment@clarin.eu

SCCTC Activities 2020-2021

CLARIN Annual Conference 2021

Martin Matthiesen



Standing Committee for CLARIN Technical Centres: Basic facts

- Ensure technical consistence between centres
- Liaise with Centre Assessment Committee
- Report to Board of Directors
- Forum to exchange ideas and experiences
- sc-centers@lists.clarin.eu
- <https://www.clarin.eu/governance/standing-committee-clarin-technical-centres>

Activities 2020-2021

- 9 meetings held
- Wikitongues as a co-operation example
- Closer collaboration with NCF
- Participation in CLARIN Resource Families Taskforce
- AAI: CLARIN Proxy discussion
- Metadata core components

Open Issues

- Discovery Service Uptake
- Streamlining of the CLARIN B certification process
- Activation/Reorganisation of Taskforces

Taskforces and Interest Groups

Metadata: Development of the CMDI Metadata model, metadata curation

Co-ordinators: Twan Goosen and Menzo Windhouwer, **Contact:** cmdi-tf@lists.clarin.eu

In Progress: Core components and FAIR vocabularies

Federated Content Search: Full Text Search across CLARIN centres

Co-ordinator: Leif-Jöran Olsson, **Contact:** tf-fcs@lists.clarin.eu

In Progress: Promoting FCS uptake

Persistent Identifiers: One cornerstone of good data management

Co-ordinators: Daan Broeder and Tero Aalto **Contact:** <https://clarineric.slack.com/messages/C60QVP5T7>

In Progress: CLARIN PID documentation

Authentication and Authorization Infrastructure (AAI): Advancing cross-border login inside and outside of CLARIN

Acting Co-ordinator: Martin Matthiesen, **Contact:** tf-aai@lists.clarin.eu

In Progress: Lindat's Attribute Aggregator uptake, CLARIN Proxy discussion

Fedora SIG: Interestgroup for users of Fedora Commons within CLARIN

Co-ordinator: Thomas Eckart, **Contact:** lg-fedora@lists.clarin.eu

In Progress: Experience sharing about Fedora 6



Knowledge Infrastructure

Bente Maegaard
Chair of KIC

CLARIN Annual Conference
29 September 2021



Technical Infrastructure and Knowledge Infrastructure

CLARIN comprises two parallel and complementary infrastructures:

- the technical infrastructure to share data and tools, and
- the knowledge infrastructure to share knowledge and expertise

Knowledge Infrastructure Committee – KIC

Our task as described in the by-laws

The KIC's roles are twofold:

- Advise the BoD about:
 - the KI strategy and policy, i.e. how to shape and implement the KI;
 - the promotion of the KI-related activities;
 - approval of applications, e.g. for the establishment of new Knowledge centres and for funding instruments for which opinion is sought.
- Implement:
 - parts of the strategy, in agreement with the BoD.

Hence, the activities of the KIC include both policy-related tasks and operational tasks.

Activities the Past Year

K-centres are the most important part of our portfolio

<https://www.clarin.eu/content/knowledge-centres>

- Two new K-centres have been approved,
 - CLARIN Knowledge Centre for Computer-Mediated Communication and Social Media Corpora (CMC)
 - CLARIN Knowledge Centre for Dutch (K-DUTCH)
- CLARIN now has 25 K-centres, and one more application is in the pipeline
- A first workshop for K-centres was organised Nov-Dec 2020. Purpose: promote collaboration between K-centres. Very well received, will be repeated.
- As a result, a group of K-centres have started a collaboration on sign language.

Activities the Past Year

+ Plans

- Work has been going on to improve the findability of K-centres through making sure the right keywords are associated with each K-centre
- KIC has 7 members, out of these KIC got 4 new members in May 2021

Plans

- Continue supporting K-centre visibility, try to make sure we have the right K-centres
- Organise K-centre workshop, investigate other ways of supporting K-centres and their collaboration
- Visibility: Unfortunately, the current structure of the CLARIN website does not show the K-centres in a prominent way, we hope this will soon change

see you @

www.clarin.eu

or

bmaegaard@hum.ku.dk;

Legal and Ethical Issues Committee (CLIC)

Pawel Kamocki
Chair of CLIC

CLARIN Annual Conference
29 September 2021



CLIC's main missions

- Advise the BoD on 'legal and ethical issues'
- Prepare and publish analyses
- Organise and participate at competency building events
- Consolidate recommendations related to legal and ethical issues
- Maintain the set of CLARIN license templates
- Publish and promote CLARIN policies.

Lineup Changes 2021

- New chair (since March 2021): Paweł Kamocki (DE)
- New vice chair (since March 2021): Vanessa Hanneschläger (AT)

- New members from LT, IT, NO
 - trend: increased participation of trained lawyers.

Activities 2021

- 2 meetings so far (frequent e-mail discussions in informal sub-committees)
- CLARIN Café (March) on rights of data subjects (80+ registered participants)
- Joint networking meeting with DARIAH ELDAH
- 2 papers for this conference (Kelli et al., Kamocki et al.)
- Chapter (Kamocki, Kelli, Lindén) for the CLARIN 10th anniversary book.

Plans (2021 and beyond)

- CLARIN Café (October) on text and data mining exceptions
- GDPR Handouts for language research and archiving of language resources (to replace the 2018 CLIC White Paper on the GDPR)
- White Paper on the use of social media data in language research.

join us at

legal@lists.clarin.eu

or write to

kamocki@ids-mannheim.de

Standards Committee

Piotr Bański
Chair of the CSC



CSC's main goals

- to collect, consolidate and prepare for publication in a single place its findings and recommendations related to standards;
- to maintain the set of standards supported by CLARIN and adapt them to new developments within or outside CLARIN;
- to publish and promote the standards supported by CLARIN;
- to develop and implement procedures for the discussion of recommendations and the adoption of new standards;
- to ensure harmonisation of standards between CLARIN ERIC and related initiatives;
- to ensure communication with international standards bodies such as (but not limited to) ISO;
- to advise the BoD in all matters related to standards.

Activities in 2020-2021

- 10 meetings held since CAC-2020
- Pursuing the statutory goals: unified recommendations for data deposition formats; arriving at version 1.0 just this month
- The SIS ([Standards Information System](#)) is there to gather it all in one place, in a way that allows the centres to easily update and modify their recommendations, and the user to easily access them in various visualizations
- Collaboration with DARIAH Guidelines-and-Standards WG initiated
- Chapter in the CLARIN 10th anniversary book (Bański and Hedeland)
- Presentation and demo at Bazaar-2021

Plans (2021 and beyond)

- Smoothen the wrinkles, visualize the long (and growing) list of recommendations better; add more format descriptions
- Invite centres to use the SIS rather than (or alongside) recommendations placed at their own home pages; find a spot in the B-centre checklist...?
- See if/how the information collected can be useful to other parts of the infrastructure (Switchboard for the collected media types, Bagman to be less agnostic about formats, etc.)
- Offer a CLARIN Café towards the end of the year or (more likely) at the beginning of 2022, and recruit new members
- Eventually: revamp the standards-related (as opposed to format-related) part of the SIS

Committee ‘visiting card’:

<https://www.clarin.eu/content/standards>

Feedback collected at

<https://github.com/clarin-eric/standards/issues>

or write to

banski in-the-domain ids-mannheim.de

User Involvement Committee

Darja Fišer
Chair of UIC

CLARIN Annual Conference
29 September 2021



Scope



- Tasks
 - facilitating the UI initiatives conducted by CLARIN ERIC
 - promoting the UI financing instruments within their consortium and networks of researchers
 - promoting UI participation at the relevant national and international UI events
 - facilitating the UI surveys conducted by CLARIN ERIC
- Chair: Darja Fišer
- Vice-chair: Maria Gavrilidou, CLARIN:EL
- Office support: Jakob Lenardič



- Tour de CLARIN Vol. IV: 1 consortium, 4 K-centres, 2 B-centres
 - Portugal
 - SAFMORIL, CORLI, CLASSLA, NLP:EL
 - **ARCHE, CLARIN-PL**



- Intense curation efforts
- Gap analysis for 2022 (6 new families)
- Qualitative best-practice guide for new deposits (in preparation)
- CRF projects: 1 approved, 2 under review
- Encouragement of depositing of existing impactful resources in CLARIN

CLARIN ERIC UI activities

- 19 public virtual workshops and training events
 - 10 CLARIN Cafés on a wide variety of topics
 - e.g., parliamentary data, legal issues, oral archives
 - SSHOCing Drama in the Cloud Webinar
- Ambassadors' programme
 - 2 appointed
 - 2 nominated
- CLARIN Monograph
 - to be published in 2022 by deGruyter
 - 32 accepted papers



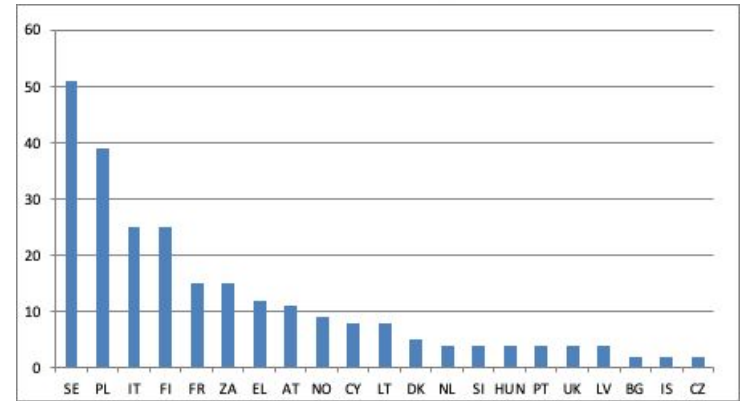
Eva Soroli



Paul Rayson

National UI activities

- 253 UI activities recorded
 - 11 fewer than in 2020
 - > 10: Sweden, Austria, Italy, Finland, France, South Africa, Greece
 - ≥ 1: all members & observers except Germany, Croatia, Estonia
 - duration: 1-day events (77%)
 - attendance: 11,068 in total
 - type: teaching events (38%), academic dissemination (30%)
- Training activities
 - mostly online due to the COVID-19 pandemic
 - e.g. course on CLARIN-PL tools @Distant Reading Training School
 - also F2F
 - e.g. CLARINO's Research Course on Data Search and Preparation
 - 1st CLARIN:EL Summer School "Language Technology and SSH"
- Media dissemination
 - written materials (e.g. FIN-CLARIN's research-of-the-month interviews)
 - YouTube videos (e.g. showcase of the Swedish National Archives on creating BERT models based of 19th cen Swedish texts)
 - FIN-CLARIN's involvement in a popular Finnish podcast on the Donate Speech campaign



- Linguistics (124) -> 62 more than in 2020
- Humanities (116) -> 27 more than in 2020
- Computer Science (66) -> 29 more than in 2020
- Literary studies (52) -> 39 more than in 2020
- Political sci & sociology (31) -> 19 more than in 2020
- History (29) -> 3 more than in 2020

- English (79)
- Swedish (39)
- Polish (35)
- Italian (22)

Committee 'visiting card':

<https://www.clarin.eu/governance/user-involvement-committee>

Questions, comments and suggestions:

darja.fiser@ff.uni-lj.si

Subscribe to UIC mailing list:

userinvolvement@lists.clarin.eu