

Title	Minutes of the 1 st User Involvement meeting at the CLARIN Annual Conference 2019
Version	1.1
Author(s)	DF, JL
Date	7–11–2019
Status	For distribution
Distribution	UI
ID	CE-2019-1531



Participants, Time, Location

Acting Chair: Darja Fišer

Attendees (in alphabetical order): Ilze Auziņa (LV), Liané van den Bergh (SA), Daniel Brodén (SE), Francesca Frontini (FR), Maria Gavriilidou (EL), Olga Gerassimenko (EE), Melanie Grumt Suárez (DE), Arjan van Hessen (NL), Barbora Hladká (CZ), Jolanta Kovalevskaite (LT), Jakob Lenardič (ERIC), Mietta Lennes (FI), Petya Osenova (BG), Eiríkur Rögnvaldsson (IS), Krešimir Šodat (HR), Jan Wiczorek (PL), Martin Wynne (UK)

Administrative support: Jakob Lenardič (minutes)

Date and Time: 2019–09–30, 11:00–13:00 CET (UTC+2)

Location: CLARIN Annual Conference 2019; Leipzig, Germany

0. Agenda

1. Welcome and introduction
2. Bylaws of the UI Committee
 - 2.1. Nomination of Chair and Vice-Chair of the UI committee
3. User Involvement Report 2019
 - 3.1. Report on CLARIN ERIC UI activities
 - 3.2. Report on National UI activities
 - 3.3. CLARIN Ambassadors
 - 3.4. Tour de CLARIN
 - 3.5. CLARIN Resource/Tool Families and showcases
4. UIC Workplan 2020
 - 4.1. UI outreach
 - 4.2. University programmes and summer schools
 - 4.3. Training materials
 - 4.4. COST actions
5. Any other business

1. Welcome and introduction

Darja F. opens the meeting and states the goals. The agenda for the meeting is approved. All the participants introduce themselves.

2. Bylaws of the UI Committee

Darja F. announces that the [bylaws](#) of the newly established UI Committee (UIC) were approved on 9 April 2019. According to the bylaws, the UIC nominates a chair and vice-chair of the committee which is submitted to the BoD for approval. Darja F. explains the roles of the chair and vice-chair:

- (i) running the UIC (mailing list, dashboard, annual meetings),
- (ii) arranging the reporting initiatives (e.g., the annual overviews of ERIC- and national-level UI events),
- (iii) facilitating the UIC initiatives (currently CLARIN Resource Families surveys and Tour de CLARIN but subject to change),
- (iv) proposing new UIC activities for the following year.

Petya O. nominates Darja F. for the role of chair. For the vice-chair, Arjan van H. volunteers and Petya O. nominates Maria G.

3. User Involvement Report 2019

3.1. Report on CLARIN ERIC UI activities

Darja F. reports that, in 2019, CLARIN ERIC organized 5 workshops, co-funded 2 summer schools and 1 hackathon, and had 1 contribution at a third-party event (full report is available [here](#)). Darja F. explains that while the number of CLARIN ERIC UI activities was lower than last year, the events were more substantial (e.g. the PARTHENOS Workshop for CEE Countries is a good example of a high-impact User Involvement event, since it is aimed at non-CLARIN countries in Central and Eastern Europe as well as facilitating cooperation with external initiatives, such as COST Action on Distant Reading).

Darja F. also reports on the results of the survey evaluating the UI events financing instrument which was well known, much appreciated and found administratively light by all the respondents. Several bottlenecks were identified: some consortia are understaffed for UI events, some consortia are limited to offering UI events in the national context. To make the call available also to such initiatives, in the next version of the call, the funding opportunity will be extended to national events in national languages with invited CLARIN experts from abroad; should the budget use reach the point of exhaustion, with the consequence that the BoD has to become more selective, priority will be given to high-quality proposals with an international perspective.

3.2. Report on National UI activities

In comparison to the previous year, the national CLARIN consortia organized 37 more UI activities (a total of 229). The higher number is especially significant since stricter criteria were followed for the inclusion of events so that only events aimed at user involvement and not e.g. academic conferences were taken into account. Darja F. also reports that the granularity of reporting in 2019 was more even across the consortia (full report is available [here](#)).

The German consortium organised the most UI events – 53 in total. Melanie G. S. clarifies that this high number is due to the fact that the German consortium constitutes 8 separate working

groups, each addressing a specific set of research topics; consequently, the German consortium has the unique possibility to address a large and diverse number of researchers.

Darja F. also highlights the general high quality of the teaching activities organised by the consortia which also produced tangible bilingual teaching materials (e.g. by the Finnish consortium).

Arjan van H. asks why the reporting form lists disciplines such as *General Humanities* and *Literary Studies*, where the latter can be understood as a subtype of the former, at the same level of reporting. Darja F. clarifies that the selection of the disciplines should reflect how the event in question is advertised with respect to the target participants; e.g., an event which focuses on online literature can be aimed at a more general humanities public or literary experts.

3.3. CLARIN Ambassadors' activities

Darja F. announces that three researchers have been appointed CLARIN ambassadors in 2019 – Francesca Frontini, Toine Peters, and Maciej Maryl. CLARIN ambassadors are established researchers who promote CLARIN at key scientific events and projects in their own research communities. Since the start of the ambassadorship programme, the ambassadors have carried out 6 contributions at scientific events, 3 contributions at project, network, institutional and research infrastructure meetings, and 1 media dissemination activity. Francesca F., briefly presents her ambassadorship experience (i.e., language resources, natural language processing and corpus stylistics) and reports that many researchers expressed an interest in using CLARIN services or collaboration with CLARIN experts, thus requiring the capacity for follow-up activities within the CLARIN network.

3.4. Tour de CLARIN

Darja F. presents the Tour de CLARIN activities that were carried out in 2019 – since the last CLARIN Annual Conference in Pisa, Italy, 7 new consortia were featured in the initiative (Estonia, Latvia, Italy, Denmark, Slovenia, Hungary, and Bulgaria). Importantly, the focus of the initiative was for the first time extended to CLARIN K-Centres – 4 centres had been covered thus far (the Centre for Treebanking, the SWELANG Centre for the Languages of Sweden, TalkBank, and Knowledge Centre for Corpus Linguistics). Darja F. also announces that the second volume of the Tour de CLARIN publication will be published by the General Assembly in November.

Darja F. expresses her gratitude to the attendees because of their tremendous help in evolving the initiative in such way that the content is now wholly provided by the consortia/K-Centres themselves, while Darja F. and Jakob L. are only involved in coordination and editing.

Olga G., who helped provide content for presenting the Estonian consortium, expresses her thanks to Darja F. and Jakob L. for their help in the editorial process. She also encourages all the consortia which have not yet been presented in the initiative to get involved.

Darja F. also reports that the consortia find the Tour de CLARIN publication very helpful for promotion in local and regional events, so the contents of Volume 1 were divided into separate brochures for use by individual consortia. The plan is to do the same for Volume 2.

3.5. CLARIN Resource Families

Darja F. announces that since the CLARIN Annual Conference 2018, 7 new overviews of CLARIN resource families were conducted for 2 corpus families (manually annotated corpora and literary corpora) and 5 families of lexical resources (lexica, dictionaries, conceptual resources, glossaries, and wordlists). By the end of 2019, the Resource Families initiative is to be expanded to include NLP tools offered by CLARIN. The first tool family to be overviewed is NLP tools for processing historical texts and will be done in collaboration with Martin W.

Darja F. expresses her gratitude to the UI representatives for their invaluable contributions to the surveys and encourages them to peruse the reports on the individual resource families that are uploaded in the CE archive to try and solve the identified issues for their respective consortia. Darja F. highlights two types of recurring metadata issues in the overviews of lexical resources:

- (i) There was a lack of a uniform reporting for size; some lexical resources reported token size, whereas others only reported the byte size of the downloadable files, which is a less useful type of information from the perspective of SSH users;
- (ii) The criteria for naming resources are inconsistent – e.g., a lexical resource is explicitly named as a lexicon, but is then described as a dictionary in the metadata.

Olga G. points out that trying to solve a metadata issue for a single resource could potentially entail a large amount of detective work; e.g., it might be difficult to get updated information on a surveyed resource if the resource is relatively old because the uploaders/original curators might no longer be affiliated with the hosting institution.

Darja F. asks the attendees for suggestions for the improvement of the methodology for resource documentation.

Maria G. suggests that the UI committee should propose a list of user-oriented tags that would be added to each resource listed in the CLARIN Resource Families. This would improve findability of the resource families on search engines like Google, since lengthy textual descriptions decrease the search ranking. Maria G. also suggests that the tags for resource families should be based on a semi-closed list, explaining that the Greek consortium CLARIN:EL currently offers a free text field where LRs providers give a description of the resource, which often does not meet the needs of the potential users and the great variety of often arbitrary descriptions hinders the retrievability of the resources.

Martin W. asks how the online webpages of the CLARIN Resource Families are kept up to date. Darja F. responds that up until the 2019 CLARIN annual conference, the updates were done manually, but from October on, a semi-automatic method will be employed where the online versions of the resource families will be automatically generated from the Google Spreadsheets in which the original overviews of the resources are presented.

Arjan van H. suggests that the VLO should include a special facet which will display whether a given resource in the VLO belongs to a CLARIN Resource Family. Darja F. responds that this is already on the to-do list by the technical developers.

Francesca F. suggests that the webpages of the CLARIN Resource Families should allow users to subscribe to an e-mail notification system which would automatically send an e-mail announcement to the subscribers when a new resource is added. Darja F. agrees with the proposal which will be forwarded to the CLARIN office.

Darja F. announces that future work in relation to the CLARIN Resource Families will focus on user-oriented showcases. The showcases will likely be based on interviews with experts who have successfully used CLARIN Resource Families corpora or lexical resources in their own work.

4. UIC workplan 2020

4.1. UI outreach

Darja F. proposes to focus on the following SSH disciplines in 2020:

- Second Language Learning
- Literary Studies
- Political Science
- Migration studies

Arjan van H. claims that CLARIAH-NL is struggling with UI outreach in the Netherlands in two respects. Members of the Dutch consortium are often unsure which language should be used in their national outreach activities, in that Dutch is more appropriate for reaching out to local scholars whereas the advantage of English is its international appeal. It is often unclear who the target audience is and consequently to what degree of precision the CLARIAH-NL activities are to be presented in the outreach events. Maria G. also reports that the CLARIN:EL consortium would be grateful for concrete guidelines for such outreach.

4.2. Universities and summer schools

Darja F. proposes that a more systematic way of integrating CLARIN content and services in the curricula of universities and summer schools should be adopted. Darja F. also announces that on 4 and 5 November, KSIC is organizing a workshop at which CLARIN members will report on existing experiences of integrating CLARIN in the classroom environment and welcomes consortia who have not yet nominated a participant to do so as soon as possible.

4.3. Training materials

Darja F. reports that the [training materials](#) produced in the framework of the PARTHENOS project were successfully demonstrated at [The Lorentz Workshop: Digital Humanities - The Perspective of Africa](#), which took place on 1–5 July 2019 and was aimed at the articulation of the specific developments in the field of digital humanities (DH) that are taking shape in Africa and their potential to enhance the global DH agenda. Darja F. explains that even though the presented use cases in the PARTHENOS training modules did not include African languages or resources, the participants successfully used the module in their own work by abstracting away from the target language of the use cases.

Darja F. proposes that national consortia disseminate their own existing training materials that were successfully used in prior UI events. Mietta L. comments that one of the problems of the

existing materials is that they are attached to an existing course which on a whole is not focused on CLARIN activities and that such materials might be unusable in a different context. Darja F. responds by claiming that such training materials would also be valuable purely as a showcase of good practice (“lessons learned”) or could even be used as templates for the preparation of future materials.

Francesca F. proposes that CLARIN should create and disseminate teaching materials that present concrete scenarios describing best practices on using CLARIN services – e.g., how to use the CLARIN Resource Families in SSH research.

Petya O. asks whether there exist promotional materials on CLARIN that national consortia could use in order to persuade external researchers or other related stakeholders (such as parliament members) to use CLARIN services. Darja F. responds that the CLARIN office provides many such materials which can be used for promotion.

4.4. COST actions

Darja F. announces that, at the [PARTHENOS Workshop for CEE countries](#), CLARIN will facilitate cooperation with researchers involved in the COST Action [“Distant Reading for European Literary History”](#).

5. Any other business

Maria G. says that the first volume of Tour de CLARIN was very useful for presenting CLARIN activities to the Greek national stakeholders, since it incentivizes them to be more proactive in their support for the Greek national consortium.

Maria G. furthermore suggests that the User Involvement dashboard should be regularly maintained.

Martin W. suggests that e-mails posted to the User Involvement mailing should be archived in a single location, possibly on the UI dashboard.

Darja F. announces that she will try to propose a new type of mobility grant in the form of a writing retreat where researchers could travel to a foreign institution in order to write and finish a paper related to CLARIN.