


Title	User Involvement Report for 2017	
Version	0.2	
Author(s)	DF	
Date	14-11-2017	
Status	final	
Distribution	GA	
ID	CE-2017-1096	

User Involvement Report for 2017

1 Introduction

This document the User Involvement Report for 2017 that is based on the User Involvement Action Plan for 2017. It is organized into 5 main sections: Section 2 – UI events, Section 3 – UI surveys, Section 4 – UI outreach, Section 5 – UI strategy.

2 UI events

2.1 UI events organized by CLARIN ERIC

In 2017, most UI events were organized within the CLARIN-PLUS project. Two of the three CLARIN-PLUS workshops relevant for UI were aimed at researchers from various disciplines working with the same type of language data and one was organized for new CLARIN member countries on how to best start UI activities. In total, 8 invited speakers presented success stories, 4 lecturers gave tutorials or demonstrations of the key tools in the field, 7 speakers from CLARIN ERIC presented the CLARIN infrastructure, 40 participants presented their work and there were 6 discussion sessions on the most important open issues, action plan and future collaboration opportunities. The talks were recorded and a selection of them has been published on the Videlectures portal (see section 4.2).

In addition to the CLARIN-PLUS workshops, we have started organizing CLARIN UI events that are co-located with major conferences and workshops in NLP, Digital Humanities and Social Sciences. In connection with the Language, Data and Knowledge conference, we have organized a CLARIN tutorial on text analytics for Digital Humanities with 3 lecturers. After the tutorial, a CLARIN Masterclass or “researcher clinic” was also given to PhD students and early-stage researchers at which researchers had the opportunity to discuss their work with the lecturers who advised on methodological issues, resources and tools.

The events were well attended and extremely well received, the interdisciplinary dimension of the programme and opportunities to establish cross-disciplinary and cross-border connections were especially highly praised by the participants. We therefore plan to continue with this type of events in the future.

2.1.1 CLARIN-PLUS WORKSHOPS

- [Working with Parliamentary Records](#), 27-29 March 2017, Sofia (Bulgaria)
 - 3 invited talks
 - 2 CLARIN ERIC talks
 - 2 hands-on tutorials
 - 20 talks
 - 3 discussion sessions
- [Creation and Use of Social Media Resources](#), 18-19 May 2017, Kaunas (Lithuania)
 - 5 invited talks
 - 1 CLARIN ERIC talk

- 1 hands-on tutorial
- 1 demo session
- 14 talks
- 1 discussion session
- [User Involvement](#), June 2017, Helsinki (Finland)
 - 4 CLARIN ERIC talks
 - 6 talks by representatives from national CLARIN consortia
 - 2 discussion sessions

2.1.2 CLARIN TUTORIAL

- [Doing text analytics for Digital Humanities and Social Sciences with CLARIN \(LDK tutorial\)](#), 18 June 2017, Galway (Ireland)
 - 3 invited lecturers
 - Master Class

2.2 UI events (co)funded by CLARIN ERIC

In addition to centrally organized UI events, we offered a model to (co)fund events and activities carried out by national consortia. The call was open throughout the year and applications were assessed on a first-come, first-served basis. In total, we have received 7 applications, 3 for summer schools, 2 for tutorials, 1 for a masterclass and 1 for a workshop, all of which were accepted. The events offered a wide range of topics and target diverse research communities which undoubtedly boost our outreach and uptake efforts beyond the capacities of CLARIN ERIC alone. Having said that, we would have liked to see more applications on an annual basis, which is why we decided to carry out an internal survey on funding models of UI activities. We will use the results of the survey to adjust the current models in order to make sure they are as relevant and accessible to the national consortia as possible.

2.2.1 Summer schools

- [Summer School “Semantic Technologies and Linguistic tools for Digital Humanities”](#), 03 July 2017 to 05 July 2017, Madrid (Spain)
- [Culture & Technology – 8th European Summer University in Digital Humanities](#), 18 July 2017 to 28 July 2017, Leipzig (Germany)
- [TransTech17 3rd Summer School in Translation Technologies](#), 04 September 2017 to 08 September 2017, Ljubljana (Slovenia)

2.2.2 Workshops

- [Workshop on Teaching NLP for Digital Humanities \(Teach4DH\)](#), 12 September 2017, Berlin (Germany)

2.2.3 Masterclasses

- [LexMC: Lexical Data Masterclass](#), 4-8 December 2017, Berlin (Germany)

2.2.4 Tutorials

- [Tutorial at CMC-Corpora 2017: 'How to use TEI for the annotation of CMC and social media resources: a practical introduction'](#), 04 October 2017, Bolzano (Italy)
- [How to Analyse Large Volumes of Online Text \(NLP Approaches with Python\)](#), 13 October 2017, Brussels (Belgium)

2.3 UI events endorsed by CLARIN ERIC

We have started to monitor the announcements of events that are highly relevant for CLARIN and seek active involvement with them. The response of the organizers has been very positive and offered good opportunities to spread the word about CLARIN with their

audience. This is why we will continue to build a databank of such events and seek active involvement in the future.

- [International Workshop: Speech corpora, text corpora: sharing practices](#), 13 November 2017 to 14 November 2017, Montpellier (France)

2.4 Invited contributions at third-party events

We have been invited to present CLARIN at 2 major summer schools in corpus linguistics and translation technologies, 3 interdisciplinary scientific workshops and 1 international conference with which we reached over 100 students and 50 researchers. The talks were very well received and resulted in some promising follow-up proposals for closer collaboration. We will continue to look for opportunities to give lectures and talks at prominent scientific and training events in digital humanities and social sciences in the future.

2.4.1 Keynote talks

- [DARIAH Day @ UZH](#), 18 December 2017, Zurich (Switzerland): CLARIN for linguists
- [5th CMC and Social Media Corpora for the Humanities Conference](#), 03 to 04 October 2017, Bolzano (Italy): CLARIN survey of CMC resources and tools
- [TransTech17 3rd Summer School in Translation Technologies](#), 04 September 2017 to 08 September 2017, Ljubljana (Slovenia): Darja Fišer: CLARIN: Infrastructural support for machine translation and translation studies
- [CLARIN-UK: Promoting Cross-disciplinary Corpus Linguistics](#), 24 July 2017, Birmingham (UK): Darja Fišer: CLARIN - a pan-European initiative to promote Language Resources and Technologies across the Disciplines
- [Lancaster Summer Schools in Corpus Linguistics and other Digital methods](#), 27 to 30 June 2017, Lancaster (UK): Darja Fišer: CLARIN: Infrastructural support for the study and use of language as social and cultural data
- [#DariahTeach Open Education Workshop](#), 22-23 March 2017, Lausanne (Switzerland): Darja Fišer: Integrating CLARIN language resources into the classroom environment

3 UI surveys

3.1 Overviews of key families of resources and tools in the CLARIN infrastructure

An important type of the surveys we have carried out are the overviews of four key families of resources and tools in the CLARIN infrastructure: parliamentary records, computer-mediated communication, parallel corpora, and newspaper corpora. The goal of these surveys was twofold: to establish what we have to offer to researchers interested in the selected resource family and to identify gaps in our infrastructure and learn how it could be improved.

- Parliamentary Records ([Report](#))
- Computer-Mediated Communication ([Report](#))
- Parallel Corpora ([Report](#))
- Newspaper Corpora ([Report](#))

3.2 Overview of key research networks for CLARIN outreach

The goal of this type of surveys was to obtain an overview of the research landscape in Digital Humanities and Social Sciences and to identify the active research communities that are relevant to CLARIN and which could benefit from our infrastructure. In the future we will

try to establish contacts and seek involvement with a select number of the key new research communities which are not yet actively involved with CLARIN.

- COST Actions ([Report](#))
- HERA Projects ([Report](#))

3.3 Overview of requests and announcements on language resources and tools on CorporaList

We have analysed the questions in the CorporaList Archive that are looking for corpora or tools as well as the announcements of new releases of corpora or tools since 1 Jan 2017 until today. The motivation for this overview is threefold: to try and determine if the resources that are looked for are available through CLARIN, to try and determine if we could increase visibility of CLARIN by designating a representative who would reply in such cases and to try and enrich our infrastructure by contacting the authors of new corpora and tools to deposit them with CLARIN. We believe it is important to monitor what types of tools and resources are in demand as well as to keep track of what's being developed, which is why we will continue to explore and expand the monitoring as well as develop a model for follow-up activities in the future.

- Report ([Spreadsheet](#))

3.4 Focus group

In order to improve and promote the use of CLARIN's infrastructure, we have conducted two focus group interviews with two groups of (potential) users: researchers in Humanities and Social Sciences, and experts in language technology. The focus groups were conducted as part of the CLARIN- PLUS workshop on parliamentary data. The results show that both groups of researchers need more guidance about the datasets, corpora, tools and services CLARIN offers, and that the best quality materials are easily available and accessible. We will be making this as a priority for our future development efforts.

- Researchers and developers working with parliamentary data ([Report](#))

3.5 Overview of national CLARIN UI activities

The overview of national CLARIN UI activities was obtained through a survey that was sent to the national UI coordinators. We are happy to report that there was a major improvement in the reporting of the fine-grained data (e.g. 21% reporting rate for the no. of participants in 2016 vs. 87.5% in 2017).

In total, 160 events and activities were carried out, which is 30 more than in the previous year. They were attended by up to 200 participants, amounting to over 5000 participants in total. While the number of events is still very uneven across the CLARIN countries due to the substantial differences in funding, staffing, and maturity of the national consortia, it is very positive to note that all members and observers except Portugal organized at least 1 UI activity in 2017. While Finland organized the most (41), additional 4 countries carried out over 10 events and activities: Poland, Germany, Austria and Sweden.

In terms of disciplines, 92 events belonged to the general or digital humanities, 82 events were in the domain of linguistics, 46 in social sciences, 27 in computer science and 26 in literature. Compared to last year, there is a significant shift from (corpus) linguistics which was the dominant discipline last year to (digital) humanities, which prevailed this year.

Only 107 reports contained information on the language. Most represented were English (30), Finnish (21), German (18), Polish (17), and Dutch (15).

Most of the reported were 1-day events (56 %). Most events and activities were of a national character. 30 activities involved participants from multiple countries. The most frequent activities were organizations of UI events (29%) and teaching activities (27%). While most of

the consortia engaged in teaching activities (14 out of 20), information events were the least frequent (reported solely by the Finnish consortium). 48% of the events were organized by the consortia, 19% were balanced cooperation between CLARIN and non-CLARIN actors and 33% were organized by external initiatives

4 UI outreach

4.1 [Tour de CLARIN](#)

“Tour de CLARIN” is a new CLARIN [ERIC](#) initiative that aims to periodically highlight prominent User Involvement (UI) activities of a particular CLARIN national consortium. The highlights include an interview with one or more prominent researchers who are using the work of national consortium’s infrastructure and can tell us more about their experience with CLARIN in general; one or more use cases that the consortium is particularly proud of and any relevant user involvement activities carried out. “Tour de CLARIN” helps to increase the visibility of the national consortia, reveal the richness of the CLARIN landscape, and to display the full range of activities throughout the network. The content is disseminated via the [CLARIN Newsflash](#), [blog posts](#) and linked to on our social media: [Twitter](#) and [Facebook](#).

- Tour de CLARIN: THE NETHERLANDS | October & November 2017
 - [CLARIAH-NL presents the MIMORE tool](#)
 - [Tour de CLARIN: The Netherlands](#)
- Tour de CLARIN: AUSTRIA | August & September 2017
 - [Interview with Dr Stephan Procházka from Austria](#)
 - [The ACDH Tool Gallery](#)
 - [Resource from Austria: ABaC:us Corpus](#)
 - [CLARIN AUSTRIA presents 'Viennese Lexicographic Editor' tool](#)
 - [Tour de CLARIN: Austria](#)
- Tour de CLARIN: SWEDEN | June & July 2017
 - [Resource from Sweden - the Riksdag's open data](#)
 - [Interview with Maria Ågren from Sweden](#)
 - [Tutorial and workshop on automatic sentence selection from corpora](#)
 - [Korp tool](#)
 - [Tour de CLARIN: Sweden](#)
- Tour de CLARIN: FINLAND | April & May 2017
 - [The Language Bank of Finland's Roadshows](#)
 - [Interview with Tommi Jantunen from Finland](#)
 - [Suomi24 resource](#)
 - [AaltoASR tool](#)
 - [Tour de CLARIN: Finland](#)

4.2 [Videlectures](#)

To increase visibility of the CLARIN infrastructure and events, we have started collaborating with the VideoLectures.NET, an award-winning free and open access educational video lectures repository. The portal, which currently contains over 23.000 videos by almost 15.000 lecturers from academic and promotional events from many scientific fields, is aimed at promoting science, exchanging ideas and fostering knowledge sharing by providing high quality didactic contents not only to the scientific community but also to the general public. CLARIN videos are synchronised with the slides, subtitling in 5 European languages is also possible. So far, content has been limited to CLARIN ERIC events only. In order to build a comprehensive CLARIN knowledge base, we plan to offer CLARIN national consortia to deposit selected videos from their most important and successful events on the portal as well.

- [6th CLARIN Annual Conference, Budapest 2017](#)
 - 2 invited talks
 - 1 presentation by local organizer
 - 12 regular talks by participants
 - 1 presentation of the Steven Krauwer Award ceremony
 - 2 interviews with invited speakers
 - 3 interviews with authors of accepted papers
 - 3 interviews with conference participants
- [Doing text analytics for Digital Humanities and Social Sciences with CLARIN \(LDK tutorial\), Galway 2017](#)
 - 2 invited talks
 - 1 collection of interviews with participants
- [CLARIN-PLUS workshop "Creation and Use of Social Media Resources", Kaunas 2017](#)
 - 5 invited talks
 - 5 regular talks by participants
 - 1 collection of interviews by participants
- [CLARIN-PLUS Workshop "Working with Parliamentary Records", Sofia 2017](#)
 - 2 invited talks
 - 8 regular talks by participants
 - 1 collection of interviews by participants
- [CLARIN-PLUS workshop "Working with Digital Collections of Newspapers", Leuven 2016](#)
 - 1 invited talk
 - 6 long talks by participants
 - 1 short talk by a participant
 - 1 collection of interviews with participants
- [CLARIN-PLUS Workshop "Exploring Spoken Word Data in Oral History Archives", Oxford 2016](#)
 - 1 invited talk
 - 6 regular talks by participants
- [5th CLARIN Annual Conference, Aix-en-Provence 2016](#)
 - 2 invited talks
 - 14 regular talks by participants
 - 1 collection of interviews with participants

4.3 Guest blog posts

In order to actively engage with the researchers in our community, we have solicited guest blog posts from them. We have received some interesting posts that have so far been limited to reflections on events they participated. In the future we plan to invite guest blog posts on other topics as well.

- ['Reflections on the CLARIN-PLUS workshop "Creation and Use of Social Media Resources'](#) by Adrien Barbaresi
- ['Empirical data in linguistics'](#) by Ana Slavec and Jakob Lenardič
- ['CLARIN-PLUS Workshop on User Involvement by FIN-CLARIN'](#) by Tero Aalto (FIN-CLARIN)
- ['Impressions of CLARIN-LT on the CLARIN-PLUS Workshop on Social Media Resources'](#) by CLARIN-LT
- ['Reflections on CLARIN workshop on Oral History in Arezzo'](#) by Bianca Pastori
- ['Reflections on CLARIN-PLUS workshop in Sofia'](#) by Martijn Kleppe

4.4 Website

On the CLARIN ERIC website, we have designed a common web space for researchers which includes a collection of training materials, case studies and expert contacts from the entire

CLARIN network that are aimed at students and researchers of all stages who are working in the fields of Digital Humanities and Social Sciences and are interested in analyzing language data and using text processing tools that are available in the CLARIN infrastructure.

The website, called CLARIN for Researchers, is organized into the following subpages:

- Initiatives for researchers
 - Presentation of key families of resources and tools
 - Guest blog posts by researchers
 - Tour de CLARIN
 - CLARIN mobility grants
 - Videlectures
 - User Involvement events
 - Digital humanities course registry
- Who we are
 - CLARIN ERIC User Involvement coordinators
 - National User Involvement coordinators

5 UI Working group

A special interest working group (3-5 people) was established with the following members:

- Darja Fišer, CLARIN ERIC
- Tanja Wissik, Austria
- Maria Gavriilidou, Greece
- Arjan J. van Hessen, Netherlands
- Ineke Schuurman, Dutch Language Union

The UI working group had 1 full-day preparatory meeting with the following tasks:

- sharing experience and know-how among national consortia
 - what works & what doesn't
 - easily usable reusable UI scenarios
 - training materials
- discussing a model for cooperation between CLARIN ERIC and the national consortia:
 - preparing the minimum standards of user involvement activities
- discussing priorities & priority target groups

Based on the meeting, the UI-related sections of the CLARIN ERIC Strategy Document were prepared.