Title	Report on the survey for the CLARIN@universities workshop
Version	1.0
Author(s)	DF, SK
Date	13-11-2019
Status	Final
Distribution	BoD, NCF, KSIC, UIC
ID	CE-2019-1552

Report on the survey CLARIN@universities workshop

1 Introduction

The survey summarized in this report was conducted between 3 September and 31 October 2019 as a pre-workshop activity for the CLARIN @ universities workshop that took place 4-5 November 2019 in Utrecht, the Netherlands. The workshop was organized by the KSI Committee in order to share the experience with the integration of CLARIN content into university programmes within the CLARIN network. The workshop participants were nominated by the National Coordinators. The survey was sent out to all the workshop participants who were asked to either fill it out themselves or ask a colleague with more teaching experience to fill it out with the goal to gain insight into the current state of affairs, structure the workshop programme and plan future steps. We were interested in collecting individual experience and opinions of the lecturers. Wherever they could provide detailed and exact information, this was greatly appreciated but approximations were welcome as well when they could not be very precise.

The survey contained 14 questions which were organized into 3 parts: 1 Current coverage of CLARIN content in your courses, 2 Reflection on your experience with including CLARIN content to your courses and 3 Future steps and suggestions.

We received feedback from 22 CLARIN member and observer countries (Austria, Bulgaria, Croatia, Czech Republic, Denmark, Finland, France, Germany, Greece, Hungary, Iceland, Italy, Latvia, Lithuania, Netherlands, Norway, Poland, Portugal, Slovenia, South Africa, Sweden, UK). In 4 cases (France, Greece, Hungary and Slovenia), 2 lecturers filled out the survey. In 3 cases (Austria, Poland, Netherlands) the lecturer gave answers for his/her teaching activities but also included experience of his/her colleague(s). In 2 cases (Czech Republic and South Africa), the consortium provided a general summary of their activities without answering the specific questions in the survey.

2 Coverage of CLARIN content in your courses

2.1 Information about the courses where CLARIN is taught

In total, information about 55 individual courses was obtained which are taught 20 different countries (Table 1) at 31 different universities (Table 2) in. While the integration of CLARIN content ranges from 2008 to 2020, CLARIN content was included into most (43, 78%) courses in the last 5 years (Table 3). While group size ranges between 5 and 150 students per year, most courses (34 or 62%) have 10-30 students, which means that CLARIN content reaches nearly 6,000 students per year (Table 4). Most courses are taught at Masters level (31 or 56%, Table 5) and belong to the disciplines of Linguistics & Language (28 or 51%, Table 6). While rare, Digital Humanities, Culture Studies and Media Studies are represented as well, as are Computational Linguistics, Information Science and Computer Science. The most frequent course names are Computational Linguistics (5) and Corpus Linguistics (3, Table 7). Most courses are either fully dedicated to CLARIN content (22 or 40%) or present CLARIN in a single session (15 or 27%, Table 8). Nearly all teaching of CLARIN content (43 or 78%) is composed of a lecture part and a practical part while some is delivered in the form of a lecture only (7 or 13%) or a practical activity only (5 or 9%, Table 9).

Country	No. of courses
Finland	7

Italy	6
Greece	6
France	4
Lithuania	3
Bulgaria	3
Netherlands	3
Norway	3
Croatia	3
Hungary	3
Denmark	2
Latvia	2
Portugal	2
Iceland	2
Slovenia	1
UK	1
Sweden	1
Poland	1
Germany	1
Austria	1
Total	**55**

Table 1. List of countries and no. of courses reported to include CLARIN content

University	No. of courses
Aristotle University of Thessaloniki	4
Ca' Foscari University of Venice	1
Eötvös Loránd University	1
Ionian University	2
New Bulgarian University	2
Pázmány Péter Catholic University	1
SOAS University of London	1
Sofia University "St. Kl. Ohridski"	1
Tampere University	1
Université de Lorraine	1
Université Paul-Valéry Montpellier 3	3
University of Bergen	3
University of Cologne	1
University of Copenhagen	2
University of Debrecen	1
University of Gothenburg	1
University of Graz	1
University of Helsinki	5
University of Iceland	2
University of Jyväskylä	1

University of Latvia	1
University of Liepāja and Ventspils University of Applied Sciences	1
University of Lisbon	2
University of Ljubljana	1
University of Pisa	2
University of Zagreb	3
Utrecht University	3
Venice International University	1
Venice International University	2
Vytautas Magnus University	3
Wrocław University of Science and Technology	1
Total	**55**

Table 2. List of universities and no. of courses reported to include CLARIN content

Taught since	No. of courses
2008	4
2010	3
2014	4
2015	13
2016	6
2017	11
2018	3
2019	10
2020	1
Total	**55**

Table 3. Year since CLARIN content has been taught in the reported courses

No. of students	No. of courses	Total no. of students
5	3	45
6	3	54
7	1	63
8	2	72
9	1	81
10	7	90
11	1	99
12	3	108
15	3	135
16	1	144
20	9	180
25	7	225
30	3	270
40	4	360
50	2	450

75	2	675
80	1	720
90	1	810
150	1	1350
Total	**55**	**5931**

Table 4. No. of students in the courses with CLARIN content

Study level	No. of courses
Masters	31
Bachelor	16
Doctoral	3
Bachelor & Masters	3
Masters & Doctoral	1
All	1
Total	**55**

Table 5. Distribution of courses over study levels

Discipline	No. of courses
Linguistics & Linguistics	28
Translation studies	6
Digital Humanities	5
Computational Linguistics	4
Film/TV studies	3
Computer Science	2
Information Science	2
Various disciplines	2
Bulgarian Studies	1
Cognitive Science	1
History	1
Total	**55**

Table 6. Distribution of courses over disciplines

Course name	No. of responses
Computational Linguistics	5
Corpus Linguistics	3
Digital Humanities	2
Literary Translation	1
Computer-Assisted Literary Translation	1
Recent advances in Linguistics	1
Computers and language	1
Lexical Resources	1
Contrastive English-Lithuanian Stylistics	1
Natural Language Analysis	1
Corpus analysis tools	1

Introduction to research methods	1
Cognitive Science 2 (Multimodality in Communication and Cognition)	1
Language Technologies	1
Corpus Linguistics and Statistical Methods	1
Linguistic technologies for information extraction	1
Data Clinic	1
Methods	1
Digital Hermeneutics	1
Phonetics in language learning and teaching	1
Digital history in the 19th and 20th century	1
Television History Online	1
Introduction to the (computational) lexicography	1
Introduction to Speech Analysis	1
Language and computers	1
Language resources	1
Language modeling	1
Digital Humanities: Web Resources, Tools and Infrastructures	1
Language resources for machine translation	1
Digital Libraries	1
Language technologies for machine translation	1
Digital Philology	1
Lexicology and Lexicography of Lithuanian Language	1
Digital Texts and Multicultural Studies	1
Linguistics in the Digital Age	1
Digital Tools for the Humanities	1
Management of Big Information Sources	1
Digital Tools in Literature and Translation	1
Methods of digital enrichment	1
Discourse Analysis	1
Natural language processing	1
ELDP Training	1
Programming and Curation	1
Film, Text, Context	1
Specialized Translation	1
German Corpus Linguistics course	1
Bulgarian Syntax in non-native environment	1
Introduction to programming	1
Total	**55**

Table 7. List of courses reported to include CLARIN content

Coverage of CLARIN	No. of courses
Full course	22
1 session	15

Part of the course	4
Part of single lesson	3
Several sessions + assignment	3
2 sessions	2
1 session + project	2
A few sessions	2
3 sessions	1
Half course	1
Total	**55**

Table 8. Coverage of CLARIN content in the courses

Type of teaching	No. of courses
Lecture + hands-on	26
Lecture	6
Lecture + assignment	3
Online course	3
Blended learning + tutorial	2
Hands-on	2
Hands-on + assignment	2
Lecture / tutorial	2
Lecture with demo, hands-on, assignment	2
Assignment	1
Blended learning	1
Lecture + demo	1
Lecture with demos, hands-on, assignment	1
Lecture, hands-on, assignment	1
Reading seminar	1
Tutorial + hands-on	1
Total	**55**

Table 9. Type of teaching of the CLARIN content

2.2 The CLARIN content taught

As can be seen in Table 10, the most frequently mentioned content taught in the courses reported fall into the category of general issues related to CLARIN (e.g., CMDI, VLO), general issues related to research infrastructures (e.g., open science, research data management, metadata standards, research data citation, legal and ethical issues) and basic methodological/theoretical topics (e.g. language data research methods, text processing and annotation). Content related to national consortia and specific tools are also frequently taught (e.g., Korp, Weblicht, ELAN, CLARIAH Media Suite, Tundra, Praat) while specific resources are considerably less frequent (e.g., HuComTech, UD Treebanks).

Content category	No. of mentions
general CLARIN	30
general RI	29

methodology/theory	28
general resources	26
national CLARIN	19
specific tool	13
general tools	9
specific resource	3
Total	**157**

Table 10. List of content categories that are taught in the courses reported

2.3 Indicate the number and profiles of other teachers you know who also teach CLARIN content at your institution or another institution you know

Participants listed an additional 58 of their colleagues from 16 countries (Table 11) and 35 universities (Table 12) who teach CLARIN content in their classes. They mostly cover disciplines such as Linguistics & languages (11), Information and computer science (6), Humanities (5) and Language engineering (4, Table 13).

University
Budapest University of Technology
Ca' Foscari University of Venice
Danish Technical University
EUR
Ionio University
IT University Copenhagen
Mannheim University
New Bulgarian University
Paul Valéry University of Montpellier
RUGroningen
Shoumen University
SOAS University of London
Sofia University
Université Paul-Valéry Montpellier 3
University of Amsterdam
University of Bergen
University of Copenhagen
University of Eastern Finland
University of Graz
University of Helsinki
University of Jyväskylä
University of Latvia
University of Ljubljana
University of Oulu
University of Parma
University of Pisa

University of Siena
University of Tampere
University of Turku
University of Vaasa
University of Vienna
University of Zagreb
Utrecht University
Vytautas Magnus University
Wrocław University of Technology

Table 11. List of countries where fellow lecturers also include CLARIN in their courses

2.4 List the most important university programmes that could integrate CLARIN content but do not (yet)

Participants listed an additional 89 programmes at 25 universities from 16 countries that could integrate CLARIN content but do not yet do so. Most of them are at Masters level but some are also Bachelors and PhD degrees. The range of disciplines is very wide which indicates a good awareness of the wide potential of CLARIN among workshop participants.

Discipline
Acoustics and Audio Technology
Anthropology
Applied Language Studies
Applied Linguistics
Applied Statistics
Art and Design
Classics
Clinical Linguistics
Cognitive Science
Computational Engineering
Computational Linguistics
Computational Sciences
Conference translation
Cultural Sciences
Cultural Studies
Data Science
Digital and Public Humanities
Digital Humanities
Educational Sciences
Educational Technology
EMLex
English Language and Culture
Foreign Languages
General and comparative linguistics
German and Media Studies

German Studies
History
History, Culture and Arts
Information Sciences
Information Studies
Information Technology
Intercultural Communication
Intercultural Teacher Education
Knowledge Extraction and Big Data Technologies
Language and literature programmes (Danish, English, French, German etc.)
Language studies (English, French, Russian, etc.)
Languages
Languages and Linguistics
Languages and Literature
Languages, Communication and Management of Educational Services in modern social, economic and technological environment
Law
Linguistic Data Sciences
Linguistics
Linguistics (Baltic, Fin-Ugric, etc.)
Logopedics
Marketing
Master FLE
Master Humanité-Numériques
Media and Culture
Media Studies
Political studies
Psychology
Romance Studies
Saami Culture
Sciences du Langage
Semiotics, Culture and Communication
Sign Language Studies
Slavistics
Slovene language
Social Policy and Welfare
Social Sciences
Sociology
Speech Communication
Theology
Theoretical Linguistics
Transcultural Communication
Translation and Interpretation

Translation Studies
Translator-editor
Table 12. List of courses that could integrate CLARIN content

3 Reflection on your experience with including CLARIN content to your courses

3.1 Summarize the curriculum/course development process

In this section we report a selection of curriculum development descriptions (unedited) in which we tried to represent consortia with different levels of maturity, courses with different degrees of CLARIN coverage and different levels of formalities and processes of the inclusion of CLARIN content into the classes.

3.1.1 Croatia

As CLARIN RI FP7 project started in 2008-01, the first information about it appeared in three above mentioned courses from academic year 2008/2009 on. At that time the general RI explanation and introduction of the eScience paradigm was important, but after 6-7 years, the focus was shifted more to the very content of CLARIN RI and how it can help HSS researchers.

3.1.2 Finland

For the Corpus Linguistics course, we started by having a few intense meetings with several experts in FIN-CLARIN. Together, we discussed the learning goals and defined the key problems and solutions that the course should ideally address in each week. We also outlined some exercises on which the students would be working (and possibly graded) during each course week.

We had previously organized a brief intensive course in FIN-CLARIN that included at least some presentations that we were able to use as material. We also met with a lecturer in English who had previously given some lectures in Corpus Linguistics, and she offered us some of her slides as a starting point. Two lecturers (I and my colleague Jussi Piitulainen) then finished designing the lectures and hands-on exercises and rewrote the slides for the entire course. I edited about 80 % of our slides. Jussi was mainly responsible for one lecture and one hands-on session, but he assisted me in preparing the others.

Creating the content for the new lecture course in Corpus Linguistics took us at least half a year or even longer. It was very time-consuming, and we had already promised to give the course during a specific study period the next spring, so we were under a lot of pressure to finish everything on time. I and Jussi gave the classroom version of the course twice, once in spring 2014 and once in the next spring. Simultaneously during the second run of the lectures, the course materials were converted into an online course which was given by me, directly after the lecture course. In practice, the conversion process took me the entire spring term, and I was not able to finish all of the videos until the next year or so. I have also made some changes to the exercises every year in order to make my grading work easier.

In the spring of 2018, I tried running the course simultaneously in Finnish and English. Translating the materials took me a couple of months – I finished some of the translations while the course was already running.

I have just finished running the online Corpus Linguistics course for the 8th time!

FIN-CLARIN is also involved in designing a new Digital Humanities and Social Sciences study module at the University of Helsinki. The new module is to start in autumn 2020 and the study options will include the three courses that are already provided by FIN-CLARIN.

3.1.3 Germany

The development process is very informal in that I am free to determine the syllabus I follow in my classes. I usually decide to devote one or two sessions to CLARIN. Sometimes I only devote half of a session or two. I usually demonstrate how language resources are available in CLARIN, how they are findable, and how search processes are organized – in this latter respect, I talk about the VLO. When I talk about language processing, I devote one session for demonstrating WebLicht with which I show how easy the use of computational methods can be.

3.1.4 Greece

After joining the Greek consortium and uploading the first resources concerning translation, particularly monolingual and multilingual corpora, I integrated in those of my courses that have a strong technological dimension the use of this corpora in order to show a. the importance of open access linguistic resources and their interest not only for research but also for practical use b. the importance of language technology and particular tools for a translation-driven exploration of these corpora in use, e.g. terminology extraction, annotation etc. c. the importance of the metadata d. the importance of FAIR data. Although I presented CLARIN in a single session, the resources of the infrastructure are used more systematically in my courses. Next year, I plan to expand the CLARIN presentation to more sessions.

3.1.5 Hungary

Each session of the Computational Linguistics course had two main parts: a lecture on theoretical issues and a hands-on practice in the second part of the session. In the lectures, I was talking about the basics of computational and corpus linguistics: how to build a corpus, how to evaluate the performance of a tool, what kind of legal issues emerge while corpus building, how to access, use and develop language processing tools, etc. During the hands-on practice, the students could try out tools and resources, writing and running Python scripts, etc. The language resources and tools we used focused on the Hungarian language, and we also used several CLARIN corpora.

3.1.6 Lithuania

For the course “Contrastive English-Lithuanian Stylistics”, the MA programme of “Applied English Linguistics”: Among other goals, the MA programme aims to:

- educate highly qualified specialists who possess comparative knowledge and assessment skills of two cultures and languages — English and Lithuanian — and are

able to independently apply their acquired knowledge in work in academic and cultural institutions, <…>;
- educate highly qualified translators who possess knowledge of various principles of English and Lithuanian text production and editing and are able to use modern term-search and automated translation memory systems.

This programme profile is particularly compatible with CLARIN's target audience which allowed to include CLARIN content in course outline in 2018, when the curriculum of the MA programme of Applied English Linguistics was renewed. The course outline was then approved by the programme committee and uploaded to programme website. In the updated course outline, the topics "CLARIN ERIC and Virtual Language Observatory" and "Language resources in CLARIN-LT repository" were added as helping to achieve one of the course outcomes "students demonstrate knowledge of finding and using electronic language resources and tools of their analysis". The information on the infrastructure, finding and using CLARIN/ CLARIN-LT language resources and tools is presented during one week (3 academic hours). Throughout the course, some case studies based on the language resources stored in CLARIN-LT are presented and discussed. Students perform various assignments and are encouraged to access the presented resources for their research.

3.1.7 Netherlands

Currently I am still in the curriculum/course development phase for the course "Film, Text, Context". This entails discussing with course coordinators and other lecturers the possible uses and applications of CLARIN content. We expect to have the content integrated in the course in the beginning of 2020. We chose to work with this course because we thought it matched the functionalities (video annotation that support structural analysis of films) and content (silent films) of the CLARIAH Media Suite particularly well. Among the course objectives is to make – among other things - a historical contextual analysis of archival films linking various source materials.

3.1.8 Poland

The main idea behind the Natural Language Analysis course is to introduce students into the basis of Natural Language Engineering and Language Technology with some elements of Computational Linguistics (they are students of Computer Science, so talking about linguistics must be done in a cautious way). The course structure follows the main steps of a typical processing pipeline. The focus is given to the Polish language and the Polish language technology. As CLARIN-PL is the main site for the LT for Polish and also the main developer of LRTs for Polish, the whole course is illustrated by examples of CLARIN-PL LRTs and services.

3.2 The biggest positive experiences for you, students, programme and discipline

In this section we list the biggest positive experience of using CLARIN content in the classroom according to the participants. We manually classified their responses into categories in order to have a better overview of the topics mentioned. Most of the key positive experiences refer to the advantages of research infrastructures in general, reuse of resources and findability of the resources. But several benefits for the infrastructure itself, boost to students' research and their positive feedback were mentioned as well.

Advantages of research infrastructures:

- The understanding of eScience paradigm in HSS, explaining how the usage of LT not just in Linguistics, but also in other disciplines (e.g. History, Literary Theory, Psychology, etc.) can help the researchers from these disciplines.
- The objective to gather together within a single platform all European resources is valuable.
- The interfaces are usable, reliable, state of the art; the students like the idea that they can access services that are hosted in countries other than France with their University identifier; the graphical representation especially for syntactic trees are very useful.
- The presentation of CLARIN infrastructure, resources and tools help the students realize a. the importance of compiling open access linguistic resources b. their diverse uses c. the easiness of use of language technology through infrastructures like CLARIN, given it does not imply code writing or deep knowledge of language tools by them. Although in translation we use tools for many years now, use of language technology in literature analysis is much more recent and students can see the importance of language technology for new projects, enhancing research and teaching.
- Students learn how to find, use and put in pipe language resources. They learn to compare similar resources and evaluate them. They acquire awareness about license and maintenance issues.
- Accessing and using corpora and language processing tools, especially for the Humanities, is a great advantage of the CLARIN infrastructure – not only for teachers and students but for everyone.
- CLARIN content raises awareness of students and staff about computational methods. Available assets favourably increase topical work in related fields.
- It created awareness of resources and tools available
- For the students / archive depositors / training participants, having the knowledge on how to prepare and archive the primary data they collected during their own projects has contributed significantly to the quality of their collections and, simultaneously, to the increase of the amount of data they archive. They feel empowered and recognize now that the responsibility for having good collections, for research and community purposes, is mainly on their side and not on the archive's side.

Reuse of resources:

- Students have used/use CLARIN corpora and tools in master theses, projects and internships with companies.
- The students at the Computer Science have started using CLARIN resources, as well as depositing them to the CLARIN.SI repository.
- In the case of the syntax course, the positive experience was the possibility to compare typologically various linguistic phenomena across languages in a structured way – including several treebanks per some languages.
- The successes have not been due to integrating CLARIN content as such. A high-level presentation of CLARIN goals and efforts is only good for the first couple of minutes in a lecture! Rather, there has been increasingly positive feedback when actual tools, functionalities and examples have become available that can be presented in order

to illustrate relevant use cases in a concrete fashion – something that the students can immediately use for their own study projects.
- The use of CLARIN resources and tools allows the demonstration of important linguistic facts and issues.
- The use of authentic materials – there is still an overall lack of a wide selection of text corpora for digital humanities studies in Swedish.
- The students get experience on real systems, with real content (not processed for an exercise).
- Working with concrete, real data after exploring the archive in some of the courses we taught at SOAS (where we also used tools for data processing and data management), helped students to connect more easily the theoretical abstract background with concrete hands-on work. And at least one student has written her master thesis on archived materials on one language area, contributing to improve and extend the research done by the depositors.

Findability of resources:
- It was very positive to search the VLO for concrete corpora and to show the power of well curated metadata.
- I have always had the impression that people at all academic levels – so students and researchers – are generally unaware about the vast numbers of resources for an overwhelming number of that are available through CLARIN, so one of the most positive experiences is whenever I help my students become aware of what's out there.
- It made us line up basic content that would be immediately useful for students dealing with linguistic data
- They like the content that is offered (streamed content, newspapers), which was hitherto difficult to get access to, also in digital form. At first sight, they are rather intimidated by the use of digital tools. We are looking into more playful ways to educate novice users in using digital tools.
- Many easily accessible resources simplify the reference list.

Benefits for the infrastructure itself:
- Hard lesson how incomprehensible and partial are our user guides, UIs and also technical documentation.
- Practical verification of the LT developed and offered by CLARIN-PL.
- Interesting test of the robustness of our processing architecture, including also efficiency tests of a parallel, distributed system under a stress of a very large number tasks coming in the same time.
- The incipient integration of CLARIN content in teaching also motivated some of the lecturers with whom we were working more closely to archive their materials and make them accessible to the research community and the general public.

Student feedback:
- The online course in corpus linguistics is working quite well and many students and researchers say it has been relevant and useful for them.
- Student who explored Fin-CLARIN said that it is paradise for researchers.

- Usually, the students find the presented information interesting and useful for the development of both academic and professional skills and like to try out the presented tools/services. To learn about a fair reusability and open access is important part of academic competence. Sometimes students are happy when they don't have to spend time building a corpus (collecting data, creating a tool…) by yourself, and could go directly to their scientific question.

Boost to students' research:

- It enabled students to develop their research

3.3 The biggest obstacles

In this section we list the biggest obstacles participants have experienced in teaching with CLARIN content according to the participants. We manually classified their responses into categories in order to have a better overview of the topics mentioned. For many languages, domains, the lack of the resources and tools available is the key obstacle. Another major bottleneck which is commonly observed in SSH courses is the gap between the skills required to use CLARIN infrastructure with students' skills. The next two important obstacles are the unfamiliarity of teachers with CLARIN infrastructure and the lack of teaching materials.

Lack of resources:

- The lack of parallel data
- Lack of suitable language resources and tools for my language prevents us to have enough illustrative and catchy examples upon which we could demonstrate the usefulness of CLARIN RI to other HSS researchers
- Lack of resources for French, in particular of corpora and NLP modules
- I would like to use tools too, but there are not many to build usage scenarios. I also need more content, to increase the selection options. In order to use CLARIN:EL on other courses, some examples are needed.
- The lack of language resources and tools for Greek
- There are missing tools and resources for the Hungarian language
- Better and larger integration between textual resources (in particular: digital scholarly editions with critical apparatus) and language resources (e.g. dictionaries, wordnets, etc.)
- Lack of suitable language resources is to some degree an obstacle. For this course ("Film, Text, Context") the students work in Dutch which is very well suited for the materials they need to analyze. For English language courses however, the Media Suite is less suitable due to its data/resources mainly being in Dutch.

Beyond the level of students' knowledge:

- The need for learning the methodology behind the search tools that come together with the treebanks
- Huge variety of tools, methods and methodologies at and above Masters level
- TEI data format is too complex for students and for tools developers as well
- Many corpora are in TEI, and there are not many tools which allow graphical exploration or analysis of TEI corpora. Some colleagues use TXM, a system that allows you to import TEI documents, to annotate them automatically and to exploit

metadata and annotation to perform searches. However TXM is not easy to install, and requires some training to be used.
- One of the biggest obstacles we face is the lack of “modern” expertise (also among linguistic experts with several years of experience in fieldwork) but also the lack of simple background knowledge regarding the importance of creating and archiving good primary data for research and community purposes. Even though the attitude is changing, making language research data accessible to the research community in general is still very biased and not seen as a major responsibility by every researcher to contribute to the improvement of linguistic research and accountability of research results
- The lack of any linguistic background of our students. Their tendency to follow some trend to easy, non-reflexive application of very data driven `international' approaches. To convince the students that in many applications knowing more about language and LT can be profitable for their applications.
- Lack of expertise/ interest of students in Humanities faced with even very simple exercises in unix
- The knowledge-level of the DH students.

Lack of teacher experience:
- Limited expertise by teacher in the field of language data
- The lecturers must learn how to use the resources
- Lack of experience/expertise in more specific research workflows
- Lack of information and examples on resources in other CLARIN nodes, no best practice/guidelines/examples how to use CLARIN LRT in classes (e.g., if I want to show English resources –which would be most important, what needs to be stressed?)
- Another obstacle is that the lecturers themselves first need to be trained, but do not get time for it. As they have high work loads, they often do not want to invest time in learning new tools. For some tools – in particular video annotation tools – it is relatively easy to train lecturers because the tools are not that complex. However, for more advanced data analysis tools that require notions of programming, it is a different matter. With regard to the latter we also in some cases encounter skepticism because such approaches support primarily quantitative methods which are not always embraced with enthusiasm in humanistic teaching and research.
- There is a potential to use lots of other tools that I am insufficiently familiar with. There is insufficient documentation of available resources and no tutorial on how to effectively use them in the classroom. Also, some resources require extensive reformatting before they can be used. Also, there is no classification in terms of level yet (resources suitable at bachelors, masters or phd).

Lack of training materials:
- Lack of time and other resources to locate or create materials for all potentially relevant technologies
- No access to pre-prepared slide presentations that can be adapted for teaching
- Time is needed to develop specific instructions and tasks on not only how to access the language resources but also to show their usefulness in research
- Lack of resources for certain languages

- Lack of examples of use in Python for both examples and technologies.

Immature infrastructure:

- Typically, students will come to with a question like "For my term paper I want to do research on gender in Russian, where do I find data for this?" and then I am just as much at a loss as the students. In some case the VLO is very useful, in other cases it is not.
- Lack of standardization that hinders reusage of resources and tools

Other:

- Classrooms equipped with computers need to be reserved so that the students could simultaneously listen to the lecture and search the discussed websites/ try out the tools, etc.
- Limited financial resources for Faculty of Humanities
- Lack of opportunities to present CLARIN content.

3.4 How satisfactory is the current level of integration of CLARIN content into university curricula at your institution?

As can be seen from Table 13, only 2 participants are satisfied with the current level of integration of CLARIN content into curricula at their universities (Finland and Poland), where the participant from Finland notes that integration is satisfactory only in the area of Linguistics and potentially Digital Humanities whereas there is still room for improvement for other fields. The participant from Latvia notes that language resources and tools are widely used, but they are not recognized as CLARIN.

Discipline	Not bad but could be better	Satisfactory
Austria	Bulgaria	Finland
Croatia	Germany	Poland
Denmark	Greece	
France	Hungary	
Iceland	Latvia	
Italy	Lithuania	
Portugal	Netherlands	
Slovenia	Norway	
UK	Sweden	

Table 13. Satisfaction with the current level of integration

3.5 How involved has your national consortium been in this integration

In most cases (9), the inclusion of CLARIN to courses is lecturer's individual initiative. 4 participants included CLARIN at their initiative but are supported by the consortium when needed. In Finland, Greece, Italy, Lithuania and Poland the integration is driven by the consortia, while France plans to leverage individual efforts into a common action soon.

Lecturer's initiative	Lecturer's initiative but supported by the consortium when needed	Driven by the consortium	Plan to leverage individual efforts into a common action

Bulgaria	Austria	Finland	France
Croatia	Germany	Greece (1 of the 2respondents)	
Denmark	Greece (one of the two respondents)	Italy	
Hungary	Slovenia	Lithuania	
Iceland		Poland	
Netherlands			
Norway			
Portugal			
UK			

Table 14. Type of involvement of national consortia in the integration of CLARIN to a course

3.6 Do similar courses exist that incorporate other research infrastructures in their courses?

In this section we list the information that the participants have provided about the inclusion of other research infrastructures into courses at their university. Some of the mentioned items are not strictly speaking research infrastructures but we still included them because they are a good representation of what participants find relevant.

- DARIAH
    - Digital Archives und Data Management (Lecture and Exercise): The infrastructure is introduced, services are tested together with the students.
    - Textmining for Digital Humanities (Lecture and Exercise): Usage of the DARIAH-DE TopicExplorer
    - DARIAH content is being used at the Faculty of Classical and New Philologies
    - DARIAH is taught at Université de Lorraine
    - in France some lecturers use Voyant, for instance those by Aurélien Berra, in Paris. You can find more information on his blog (https://classnum.hypotheses.org/author/philologia)
    - Elena Pierazzo in Tours may use the DARIAH teach infrastructure for which she contributed an introduction to TEI
- CESSDA and ESS
    - My institution is the coordinator of CLARIN national consortium, but it is also a coordinator of two more RIs from Social Sciences: ESS and CESSDA. The colleagues sociologists integrate the information about their relevant RIs in their curricula.
- KB Delpher (newspapers)
    - There are many courses making use of KB Delpher (newspapers), that will be part of the CLARIAH infrastructure (WP6) and which have also been embedded and made searchable in the Media Suite.
- Media Ecology Project and Cinemetrics
    - Eric Hoyt's Media Lantern Project, Mark Williams' Media Ecology Project and Yuri Tsivian's Cinemetrics are used in other courses. With CLARIAH WP5 we

are looking into collaborations with the two first-mentioned projects and have already to some extent exchanged software and data.

- Korp and Sparv
    - Swe-Clarin/Språkbanken Text also uses a range of tools, such as Korp and Sparv, in teaching situations in several other courses in Swedish.

4 Future steps and suggestions

4.1 What are the most important next steps needed to increase the integration?

In this section we list the participants' responses on the key next steps to increase the integration which we manually organized into categories in order to have a better overview of the topics mentioned. Most focus on the development, exchange of teaching materials and teacher training as well as increased dissemination activities among teachers.

- Teaching materials
    - Provide teaching material and tutorials.
    - Produce training material in French.
    - If CLARIN could gather a collection of good teaching practices, national consortia and local lecturers could easily adapt and translate them, and organize presentations for the colleagues.
    - Providing teaching materials that are easy to use, adaptable, and can be used to "teach the teachers"
    - my colleagues need more examples
    - Some sample courses and course materials would be useful.
    - Preparation of resource materials for students of different profiles. E.g., a resource book with practical activities for which trainee translators have to use resources and tools stored in national CLARIN repositories as well as VLO.
    - The development of course exercises and blended learning programs that offer clear-cut examples for students and instructors of how to meaningfully analyze materials. These exercises should be small and available in an easy form (e.g. in a word document).
    - identify suitable tools for demos in classrooms, marked at different levels (beginner, intermediate, expert, corresponding to bachelors, masters, phd). Distribute presentations that teachers can use.
    - make reproducible "recipes" that teachers can demo and students can follow (e.g. find resource in VLO, download, extract selected data, process with tool, etc.).
    - ideally, students at masters and phd levels should get "recipes" that empower them to reproduce selected research results in the literature.
    - provide a list of possible topics for assignments, with pointers to resources and tools, and additional scripts for reformatting, statistics, etc. tailored to the resources.
    - Better materials for users – students on all levels, including examples how to approach different problems.
    - Some examples of the use of the infrastructure for teaching and dissemination purposes

    - Python examples of use, links to open source code
- Training
    - the key to increase the integration of CLARIN content into teaching and research is more training. Lecturers who are not directly involved with CLARIN and its activities, are not prepared to integrate and do not feel comfortable in integrating existing resources into their teaching. On the other side, it is not easy for them to get proper training in this domain. A proof thereof is the fact that during the last 3 years, we at ELDP and ELAR have been approached by several researchers and students to attend our trainings (which are normally closed to ELDP grantees), by explaining that no such training is available at their home institutions.
    - organize training events at national level
    - approaching potential teachers from diverse fields and telling them how easy it is to integrate CLARIN aspects in teaching and curricula.
    - The most important initiatives are seminars, workshops, conferences that will show the importance of language technology for different disciplines, which currently don't use linguistic resources and tools. From my experience, when colleagues experience the use of LRs and tools through practical, hands-on workshops, they want to know more and are keen to use them. This was the case with the political scientists in Aristotle University of Thessaloniki and the National Centre for Social Research of Greece.
    - I suggest the model of SunoikisisDC – An International Consortium of Digital Classics Programs, directed by Monica Berti, in which students and teachers around the world attend seminars in teleconference, and all the recorded materials are available online (e.g. https://github.com/SunoikisisDC/SunoikisisDC-2018-2019/wiki/Summer-2019:-DC). Also Massive Open Online Courses (MOOC) are a viable solution. Anyway, all educational materials should be available online and FAIR.
    - I would like to see is programs dedicated to teaching the teachers about CLARIN and thereby spread the word that the infrastructure is something my colleagues could integrated in their syllabi.
- Dissemination
    - More promotion of CLARIN content and tools is needed, so that researchers, lecturers and students are more aware of its existence.
    - Promote CLARIN services to lecturers of courses that do not deal primarily with digital topics.
    - disseminate more the CLARIN resources and tools within and outside consortium
    - The visibility of CLARIN should be improved among teachers in Humanities and Social Sciences.
    - Raise the awareness of further teaching staff.
    - It is important to approach the teachers in subjects like history, literature etc. and explain to them how they can benefit from CLARIN resources and tools. This is on my agenda as National Coordinator.
    - That colleagues would be aware of CLARIN and – more daunting – that they would be willing to change their course to incorporate CLARIN-related subjects.

- Technical development
    - To be used by language and literature teachers we need easy to use interfaces and visualization tools and disseminating the infrastructure with workshops, small meetings with lecturers from the various disciplines.
    - Encouraging the community-development of quality open-source tools with added value, which exploit CLARIN resources. To do so, we need clear and open CLARIN APIs, and CLARIN may encourage such community developments by proposing tutorials, labels, funding, hall of fame of the best community tools, etc. Hence, hackatons and practical sessions in class could focus on reusing or implementing such tools, adapted to each University.
    - Improve coverage for French
    - I need more tools and content.
    - Make PORTULAN CLARIN fully operational
    - less restricted licences of certain resources
- Ambassadors
    - higher CLARIN visibility could be achieved in different sciences via ambassadors of CLARIN who should be educated and experienced enough in the usage of CLARIN RI so they can spread the information about its usefulness. Their role would be to “evangelise” their own scientific field and make its researchers aware of CLARIN. Only then we can start seeing the CLARIN content appearing in the curricula of other HSS instead only in Linguistics and maybe occasionally in some philologies.
- Collaboration
    - closer cooperation among consortium partners
- Plans
    - The use of linguistic resources and tools used to be a priority for specific disciplines, as Linguistics, Translation Studies and sometimes Education. Nowadays, other disciplines, as the Political Sciences or Media Studies incorporate these resources and tools. Although this development is relatively slow, we plan workshops involving colleagues and students from these departments, in order to show and explain to them the importance of infrastructure resources and tools for their work, and help them incorporate CLARIN in their courses. As regards Media Studies, we plan a workshop focusing on a corpus of advertisements and its exploration. Regarding Political sciences, we plan a workshop on computer-assisted text analysis.
    - In the DH programme at GU, there is a clear plan to increase the use of LT and the Swe-CLARIN/CLARIN resources in the near future.
    - For us at ELAR, the next step will be to work on the transformation of our digital collections into corpora and improve and expand its usability in research and teaching.

4.2 How could CLARIN ERIC further support you to facilitate such integration

In this section we list the participants’ responses on how CLARIN ERIC could support such integration which we have organized into categories in order to have a better overview of the topics mentioned. Again, most focus on the development and exchange of teaching materials, and the provision of teacher training and support for teacher exchange.

- Provide materials for teachers:
    - Provide materials to get a quick overview and understanding of a topic (e.g. CMDI, Weblicht). For example, materials similar to the primers of the W3C but also simple use cases and examples. These should be easily findable through the CLARIN website.
    - Provide guidance on which datasets are particularly suitable for specific teaching purposes.
    - Provide ready-made modules on CLARIN content that could be easily incorporated into ongoing courses would be helpful. These can be low-level and address students of basic/beginner's courses. E.g. a module about the VLO and how to search it.
    - Provide case stories for lecturers in e.g. literature.
    - Different "learning blocks" could, e.g., be flexibly tagged with keywords (by anyone!) so that it would be easy for other teachers to find, understand and reuse the materials related with each topic area.
    - A list of usage scenarios and ideas. Possibly a library with "good practices".
    - Teaching materials would be another good idea, as we could have insights for material in our language and exchange ideas on them and their use
    - developing teacher material into which using CLARIN resources and tools is integrated
    - Jupyter Notebooks with interactive educational materials should be metadated, maintained and shared
    - Some sample courses and course materials would be useful.
    - Propose an initiative of selecting and publishing the best teaching practices for different target audiences (translators, linguists, historians, sociologists, etc.). Maybe this could be a sub-theme/ sub-session of the next CLARIN conference – teaching applications of CLARIN content.
    - CLARIN would have a library of teaching materials under a CC licence that could be reused/adapted/translated for local curricula or even for self-teaching.
    - creation of for instance an e-learning course on the possibilities of integration of CLARIN content and tools into teaching with concrete examples, creation of more user-friendly teaching materials
- Support the exchange of teaching materials:
    - exchanging teaching material
    - teaching material exchange
    - It should be easy for all CLARIN members to provide (links to) their materials and to be automatically cited for sharing their content (given that the default license is something like CC BY). For a teacher who creates and shares materials, it should be hassle-free to update or add new content (preferably no separate uploads if the material is already available elsewhere).
    - Exchange of teaching materials
    - teaching material exchange
    - Teaching material exchange would be very nice. Perhaps a platform for exhanging materials may be a good way to facilitate this.
    - exchange of teaching materials and examples for other LRTs or pipelines would be very useful

    - exchange of teaching material would be very welcome. Also, exchange of experiences that were successful (or not) and recommendations.
- Teacher training:
    - organizing workshops for lecturers in e.g. literature
    - dissemination workshops where we invite and evangelize university teachers who teach Digital Humanities courses but who have not yet included CLARIN content
    - organising joint workshops
    - organize courses for the teachers
    - Support more target-oriented training
- Teacher exchange:
    - Teacher exchange would be welcome for some colleagues and some courses, however the language barrier might be an issue for a certain number of them in France.
    - Facilitating teacher exchange
    - facilitate teacher exchanges, as both teachers and students could profit of others experience and expand their research horizons
    - Facilitating teacher exchange
    - teacher exchange would no doubt be beneficial
    - At the last UI meeting, it was important, to hear that CLARIN mobility grants could be used to facilitate teacher exchange, to develop CLARIN-oriented curricula.
    - Exchange of staff
    - facilitate teacher exchange
- Joint projects:
    - collaboration in transnational projects, not necessarily big ones, either by specific calls or by regularly informing the partners for collaboration opportunities

4.3 How could your national consortium further support you

In this section we list the participants' responses on how national consortia could support such integration which we have organized into categories in order to have a better overview of the topics mentioned. Again, most focus on teacher training but there are also suggestions to improve documentation and dissemination activities.

- Training:
    - Huma-Num has the capacity of organizing great webinars and training events, once their CLARIN activities are fully operational
    - there should be more than one teacher working in and on each course
    - roadshows, not only in the country but also in diasporas
    - continue organizing seminars in the Universities, in order to disseminate best practices and promote the use and re-use of CLARIN resources
    - organize workshops for lecturers who might not always be researchers or may not be involved in digital humanities
    - Organize courses for the teachers. But this may be less effective at national level than at European level because in smaller countries, the target audience

is small. Also, I'm not sure there are sufficient incentives for teachers to participate.
    - have a person who would, as part of their duties, take care of CLARIN outreach
    - By organizing more target-oriented training.
- Documentation:
    - provide more structured description on the availability of the various resources and tools
    - development of user guides, technical documentation and teaching materials
    - develop case scenarios to implement in our teaching experience. For each level / type of activity, establish a list of resources and tools available.
- Dissemination:
    - being present at specialist conferences and inform participants about CLARIN and its offerings when it comes to teaching
    - advertise new material on relevant mailing lists
- Resources and tools:
    - integration of open source tools developed by third parties
    - common projects
- Ambassadors:
    - Ambassadors should be selected by national CLARIN consortia for each of the HSSs and then they should go through the centrally organized courses that would help them spread the word about CLARIN in their own research community. This will enable not only higher visibility of CLARIN RI, but also potentially spread the community of possible CLARIN users.
- Funding
    - Provide Teaching Pilot funding to lecturers. Without any financial support (to decrease teaching load) it is difficult for lecturers to learn new tools.
- Collaboration:
    - cooperation between different universities in the consortium and the university library

4.4 What could the CLARIN community learn and adopt from other research infrastructures in this respect?

In this section we list the participants' responses on what CLARIN could learn from other research infrastructures on this topics. We have again organized the responses into categories in order to have a better overview of the topics mentioned. Apart from the more general suggestions, most mention DARIAH and other DH activities. Some of the mentioned items are not strictly speaking research infrastructures but we still included them because they are a good representation of what participants find relevant.

- General:
    - how other sciences with longer existing RIs (e.g. Physics, Astronomy, …) integrate information about their RIs in their curricula
    - would be interesting to see how far CLARIN manages to include the activities of other research infrastructures in university curricula; that is, it would be great to get in touch with researchers working in different fields – like physicists and medical researchers – and who have already successfully

integrated contents related to their research infrastructures in their curricula. We could ask them for blueprints to ease the integration.
    - making good system usage documentation, and providing more tools (that can be combined into some scenarios) and content
    - more visibility of actions undertaken and running projects and the enhancement of collaboration in supranational projects
    - learn from practices of various digital humanities in organising summer workshops involving students across disciplines
    - examples of courses, things that work well in teaching
    - how to involve the Humanities (apart from Linguistics) in the infrastructure
    - More data sharing and data discoverability as key for interdisciplinary research and innovation in the humanities.
- DARIAH:
    - 3-5-day schools or workshops on specific topics, a good way to convey topics in a concentrated form without having to integrate them in a curriculum. This compressed form of teaching often is much more useful when it comes to practical/hands-on content than a course that runs for an entire semester.
    - the more creative ideas about interdisciplinary topics for working groups and thus – topics
    - DARAH teach infrastructure is also a good example; it would be useful to have ready-to-use tutorials in English, which can be subtitled in different languages
- SSK:
    - get some inspiration from the SSK (http://ssk.huma-num.fr) and adapt it for teaching scenarios
- Europeana:
    - a nice funding scheme for researchers which may be interesting to take inspiration from in terms of seeing how content and tools may be integrated into teaching and research.
- Ranke 2.0 platform of the Luxembourg Centre for Contemporary and Digital History:
    - may offer inspiring examples of tutorials and/teaching materials for the CLARIN community team

5 Results of the workshop CLARIN @ universities workshop

More information about the workshop and the programme can be found on the workshop page: https://www.clarin.eu/event/2019/workshop-clarinuniversities.

The full set of slides presented at the workshop can be found on: https://www.clarin.eu/sites/default/files/clarinuni-slides.pdf.

Here we repeat from the slides the next steps that could be taken or recommendations to be followed de-centrally by the CLARIN National Consortia and centrally by CLARIN ERIC, and some general recommendations and open questions.

5.1 Next steps and recommendations for National Consortia

Underlined are those where CLARIN ERIC can help.

- Understand the educational landscape in your country
    - identify key courses & teachers
    - build a contact list & network of teachers
    - promote & support exchange of experience and materials among teachers
    - support teacher exchange, both nationally and internationally
    - if possible, be part of new curricula development
- Organise training activities for teachers
    - discipline-specific works best
    - training for early-career teachers works best
    - share experience with other NCs as examples
- Provide user guides and training materials
    - improve documentation and user guides for tools, resources and services & share with other NCs
    - develop new course material for your flagship resources and tools and show how to teach it
- Dissemination
    - be visible at all major key national academic events
    - distribute material nationally & internationally (as examples, or as localisable material)
    - collect appealing use cases & share with other NCs

5.2 Next steps and recommendations for CLARIN ERIC

- Technical development
    - add a facet to VLO for resources and tools that are particularly suited to be used in teaching
    - develop a facility to exchange teaching materials
- Development of teaching materials
    - develop teaching materials for general topics on RIs & central services
    - provide central support to prepare training activities for broader cross-border and cross-disciplinary use (e.g. beefing them up, making recordings, subtitling)
    - use mobility funds for collaboration on developing courses or course material
- Development of expertise
    - organize training activities for teachers & disseminate info about training activities offered by national consortia
    - add a track to report on teachers' experience in using CLARIN in their classes at the CLARIN Annual Conference
- Strategic activities
    - appoint a CLARIN Ambassador to promote the educational aspects of CLARIN
    - stimulate submissions for Erasmus+ and Marie Skodlovska ITN calls

5.3 Survey follow-up

The survey was a pilot, should there be a next one?

- Pro: we might still find new useful information or ideas
- Con: not much new to be expected from small countries, too much work in large countries, especially if we go beyond the CLARIN sites
- Suggestions:
    - National consortia carry out the survey as required/feasible

- Explain better what CLARIN content is
- Only do it if we have new questions, such as
    - Use of existing e-learning platforms (or should we develop our own)
    - Asking requirements instead of status quo

5.4 Recommendations for teachers

General recommendations for the development of courses

- Try to use CLARIN material in your courses (and mark it as such)
- Piecemeal, next step only if previous step has been properly digested and worked with hands-on on relevant and authentic questions/tasks
- Make it manageable and don't underestimate the effort required for the development of teaching materials
- Realize that for every new element in the curriculum something (and often even someone) else has to go out

5.5 Open questions and caveats

For all material collections:

- Quality control hard to organise (but maybe not necessary if audience is experienced teachers)
- Obsolescence

Integrate CLARIN in general Digital literacy courses/Research methods corpuses or in specific courses?

- Option1:
    - Pro: Can be generic and easily shareable
    - Pro: Help preparing the grounds for real digital scholarship if organised at a very early stage in the curriculum
    - Con: Not problem driven and therefore less appealing for humanities scholars
- Option2:
    - Pro: Good bypass if option1 can't be achieved for whichever reason
    - Pro: Contextualized repetition of specific issues could be a good thing for the students when coupled to a relevant problem driven teaching unit
    - Con: Only limited attention to selected topics possible, cannot be comprehensive

6 Concluding remarks

The survey and the discussion of its results in the workshop have provided a wealth of information and ideas, worth sharing with the CLARIN community as a whole.
The next steps and recommendations emerging from the workshop can be taken on board by the National Consortia, and by CLARIN ERIC.
We recommend the National Coordinator's Forum takes them into account when preparing the workplan for the next period. In addition it will be the task of the Knowledge Sharing Infrastructure Committee to monitor progress, to identify and share what works and what doesn't work, and to use the information collected as a starting point to develop a longer term strategy for better penetration of CLARIN and its offerings in university curricula.