CLARIN-D Report R8.4

Hannah Kermes, Jörg Knappen,
José Manuel Martínez Martínez, Elke Teich, Mihaela Vela

May 2015

Contents

1 Introduction

CLARIN-D continuously seeks to increase the dialogue with the prospective users of its services attempting to reach out in the different scientific communities. To this respect the major event of workpackage *training and education* in the fourth project year was the summer school which took place from July 22 to August 1st in Leipzig as a joint event with the well established *European summer university Digital Humanities Culture & Technology*. CLARIN-D co-organized the summer university, organized five workshops and provided insight into its resources and tools with a daily information desk and demonstrations. A more detailed report on the summer university can be found in chapter 2 and on the website of the summer university [1].

After the success of the first event, the preparation for the second CLARIN-D summer school has already started. The summer school will again be organized as a joint event with the *European summer university Digital Humanities Culture & Technology* and will take place from July 28th to August 7th 2015. CLARIN-D will organize four workshops and a demonstration session at the beginning of the summer university. More information about the status of the summer university can be found in chapter 3 and on the website of the summer university [2].

Furthermore, CLARIN-D continued its series of training events for doctoral students, the CLARIN-D PhD-days. The last workshop in the series *Sprachdatenbanken – von der Aufnahme zur Publikation* was held in Munich on April 9th, 2015 and was organized by the CLARIN-D center of the LMU, Munich. The next event in the series is scheduled for October 2015 (exact date to be announced) and will be organized by the CLARIN-D center of the University of Leipzig, with a strong focus on machine learning applications.
CLARIN-D also continued its rich offer of workshops, tutorials, courses, demonstrations and presentations. An overview of the training measures organized by CLARIN-D centers can be found in chapter 4.

After an internal testing period, TeLeMaCo (a collaborative portal for training and teaching materials) went public in September 2013 and was announced at the GSCL 2013 [Amoia et al.(2013)]. Since then, a steady trickle of descriptions was added to TeLeMaCo, now holding a total of 134 materials. Most of the contributions still come from members of the CLARIN-D project, but we start seeing submissions from other places, too. More information on TeLeMaCo can be found in chapter 5.

[1] http://www.culingtec.uni-leipzig.de/ESU_C_T/
[2] http://www.culingtec.uni-leipzig.de/ESU_C_T/

2 Joint CLARIN-D and European Summer University Digital Humanities Culture & Technology

The first joint *CLARIN-D and European Summer University in Digital Humanities Culture & Technology* took place from July 22nd to August 1st 2014 at the *Geisteswissenschaftliches Zentrum* of the Universität Leipzig with 61 participants from 24 countries. You can find more information on the general concept and the mission of the summer university on the website http://www.culingtec.uni-leipzig.de/ESU_C_T/ and in last-year's project report R8.3.

The program was a mixture of 12 small one or two week workshops (up to 10 participants) and plenary events such as lectures, presentations and poster sessions as well as a round table. The social programs including excursions and communal dinners offered the possibility for additional discussions and networking. Please consult the website for a complete schedule of the summer university http://www.culingtec.uni-leipzig.de/ESU_C_T/node/454.

CLARIN-D organized five workshops, four of them had CLARIN-D center members as workshop leaders.

- Andreas Witt (Institut für Deutsche Sprache, Mannheim, Germany): *Query in Text Corpora*
- Peter Fankhauser (Institut für Deutsche Sprache, Mannheim, Germany) / Hannah Kermes / Elke Teich (Universität des Saarlandes, Germany): *Comparing Corpora*
- Susanne Haaf / Christian Thomas (Berlin-Brandenburgische Akademie der Wissenschaften): *Historical Text Corpora for the Humanities and Social Sciences. Digitization, Annotation, Quality Assurance and Analysis*
- Christoph Draxler (LMU München, Germany) / Timm Lehmberg (Universität Hamburg, Germany): *Spoken Language*
- Matthias Lang / Dieta Svoboda (University of Tübingen, Germany): *Space - Time - Object: Digital methods in archaeology*

In the beginning some of the workshop leaders were skeptical, because of the diverse backgrounds of the participants. However, after the summer school all agreed that their skepticism was needless. The diverse backgrounds were not an obstacle but an enrichment, although sometimes also a challenge. The small workshop size (between 5 and 10 participants), and the fact that participants and experts were together for at least one week made it possible to discuss (individual) questions and problems in detail and to work together in order to find solutions. This was also reflected in the evaluation of the workshops by the participants ("genügend Raum + Bereitschaft Fragen zu beantworten, Projekte zu diskutieren; nette Atmosphäre", "Möglichkeit, Rückfragen zu stellen/ Ideen, welche Tools es gibt"). A complete list of the workshops can be found on the website of the summer university http://www.culingtec.uni-leipzig.de/ESU_C_T/node/371

CLARIN-D additionally organized an information desk which was open during coffee and lunch breaks with daily changing demonstrations:

- *Historical Corpora, Standards, Quality Assurance* (Susanne Haaf and Christian Thomas, BBAW)
- *Information on Legal Issues* (Erik Ketzan, IdS)
- *WebLicht and other tools* (Thorsten Trippel, Universität Tübingen)
- *Webservices and text corpora* (Volker Boehlke, Universität Leipzig)
- *Spoken Language Corpora* (Timm Lehmberg, HZSK)
- *Web Mouse and other tools* (Christoph Draxler, BAS)
- *Virtual Language Observatory* (Kees Jan van de Looij, MPI Nijmegen)
- *Adaptable Tools for Information Extraction and Visualization* (Kerstin Eckart, IMS)
- *Multilingual and Register Corpora* (Hannah Kermes und Elke Teich, Universität des Saarlandes)

Participants could get an insight into resources and technologies the CLARIN-D infrastructure offers and could discuss individual problems with experts from the CLARIN-D centers.

The summer university provided a unique possibility to communicate the possibilities provided by the services CLARIN-D offers to prospect users from different communities within the (Digital) Humanities. CLARIN-D learned more about the (individual) problems and needs of the researchers, and identified a great number of potential obstacles as well as prejudices researchers might have. We often heard sentences such as: "If only tools X or resource Y existed, it would be so useful for my research", "If ... was possible, then ..." or "Neither do I get any funding for the compilation of a resource, nor is it useful for my academic career". Specific resources are still missing in the humanities, however, the benefit of such resources is not widely agreed on, and resource building is still not considered a scientific tasks in all fields. Furthermore, it turned out that users often do not realize the full potential of existing resources and tools, as well as the need to move beyond the well-established methodologies and research questions.

As a conclusion, we can say that the summer school is a very specific, very intense, and worthwhile event, not only for the participants but also for CLARIN-D. We took every opportunity to communicate the services provided by the CLARIN-D infrastructure and attracted much interest; especially among the young researchers, reaching exactly the audience that we wanted to reach. Even though there is still a long way to go, we already see that the boundaries between disciplines become more and more transparent, paving the way for a future interdisciplinary research. We have to learn from one another to discover new methods, and to gain new insights or as a participants from a historical background stated as one important thing she learned at the summer university: "Linguists can do a lot of stuff!"

3 Second Joint CLARIN-D European Summer University in Digital Humanities Culture & Technology

The second joint *CLARIN-D European Summer University in Digital Humanities Culture & Technology* will take place from July 28th to August 7th at the *Geisteswissenschaftliches Zentrum*, Leipzig. We decided to prepone the second summer school: first, not to loose the contacts that we gained, second, to stay in the minds of the community and third, because the summer 2016 would already be after the project phase. The structure of the summer university is similar to that of its predecessor. Centered around a selection of small workshops (up to 10 participants), there will be plenaries including lectures, presentations, poster sessions, and a round table. Participant numbers are aimed at around 60 coming from all over Europe and beyond. CLARIN-D will organize four workshops, three of which have CLARIN-D center members as workshop leaders:

- Axel Herold (Berlin-Brandenburgische Akademie der Wissenschaften, Germany) / Erhard Hinrichs (University of Tübingen, Germany) / Thorsten Trippel (University of Tübingen, Germany): *Methods and Tools for the Corpus Annotation of Historical and Contemporary Written Texts*

- Peter Fankhauser (Institut für Deutsche Sprache, Mannheim, Germany) / Hannah Kermes (Saarland University, Germany) / Elke Teich (Saarland University, Germany): *Comparing Corpora*

- Laszlo Hunyadi (Debreceni Egyetem / University of Debrecen, Hungary) / Tim Lemberg (University of Hamburg, Germany) / Uwe Reichel (Institute of Phonetics and Speech Processing, University of Munich, Germany): *Spoken Language and Multimodal Corpora*

- Matthias Lang / Dieta Svoboda (University of Tübingen, Germany): *Spatial Analysis in the Humanities*

Please visit the website of the summer university for a full list of workshops http://www.culingtec.uni-leipzig.de/ESU_C_T/node/481. As a substitute for the CLARIN-D information desk, we plan to organize a demonstration session as a plenary event in the afternoon. The session will preferably take place during the first two days of the summer university. This will make the participants aware of the CLARIN-D services at the beginning and leave those interested ample time to get into contact with CLARIN-D members in situ. Aside of financing parts of the costs of the summer school, CLARIN-D will also give out scholarships covering tuition fees. We hope to be able to strengthen the connections and further increase the communication established during the first Joint *CLARIN-D Summer University in Digital Humanities Culture & Technology* so that CLARIN-D can reach out further into this community.

4 Training Modules and Tutorials

4.1 CLARIN-D Doktorandentage (PhD days)

Workshop: Sprachdatenbanken – von der Aufnahme zur Publikation
Christoph Draxler, Thomas Kisler, Bernhard Jackl, Stefanie Pletzer, Florian Schiel (LMU Munich, Germany)
http://clarin.phonetik.uni-muenchen.de/workshop
April 9, 2015, Munich

There will be another workshop in the series of CLARIN-D Dotkorandentage in October 2015 in Leipzig focusing on machine learning.

4.2 Workshops

Workshop: Vierter DTA-Workshop: Aufbau historischer Sprachressourcen: Arbeiten mit den Angeboten des Deutschen Textarchivs
Matthias Boenig, Susanne Haaf, Christian Thomas, Frank Wiegand
July 7, 2014, Berlin (BBAW)
http://www.deutschestextarchiv.de/veranstaltungen/dtaworkshop4
On July 7th the DTA (Deutsche Textarchiv) organized the workshop "Aufbau historischer Sprachressourcen: Arbeiten mit den Angeboten des Deutschen Textarchivs". The workshop addressed users of the DTA as well as others, who were interested to learn more about the services provided by the DTA. Well-known as well as new methods and tools for the compilation and analysis of corpora were presented. Hands-on experiences exemplified possible applications.

Workshop: Using CLARIN for Digital Research
Presenter: Martin Wynne (Oxford University), Thorsten Trippel (Universität Tübingen), Christoph Draxler (LMU Munich, Germany)
Digital Humanities 2014 Workshop # 012
July 7, 2014, Lausanne
http://dh2014.org/
In this workshop, several ways to use CLARIN for research were discussed. Topics included "finding resources in CLARIN", "archiving and curating resources", "increasing research impact by using CLARIN services", "integrating tools into the infrastructure" and "using CLARIN to do research".

Workshop: Historical Text Corpora for the Humanities and Social Sciences. Digitization, Annotation, Quality Assurance and Analysis
Susanne Haaf, Christian Thomas
Joint ESU DG K & T and CLARIN-D Summer University
July 22–26, 2014, Leipzig
http://www.culingtec.uni-leipzig.de/ESU_C_T/node/378

This course gave an overview of the methods for creating standard conformant and interoperable resources of historical texts. It was shown how new and existing textual resources can be built up or further processed, respectively, in order to meet the requirements of CLARIN. In this context, different possible workflows for the integration of textual resources into the infrastructure of CLARIN-D by example of the Deutsches Textarchiv project were presented. Problem fields discussed here included:

- the acquisition and provision of high quality image sources,
- guidelines for transcriptions true to the source material,
- text structuring and metadata recording according to the TEI-P5 based DTA "Base Format" (DTABf), the best practice format for the annotation of historical written corpora in CLARIN-D,
- quality assurance within the digitization process.

Several tools and services provided by the DTA and by CLARIN to support the different tasks in the process of building up CLARIN-conformant resources were presented. Finally, it was demonstrated how resources which are built up and structured homogeneously according to the DTABf may be linguistically analyzed with automatic methods and automatically converted into other standardized formats, how they may be made available and provided in the long term within the CLARIN infrastructure and how they can be analyzed by usage of CLARIN-/DTA-tools.

Course: Query in Text Corpora
Andreas Witt (Institute for the German Language, Mannheim)
July 22 – August 1, 2014
European Summer School in Digital Humanities, "Culture & Technology"
http://www.culingtec.uni-leipzig.de/ESU_C_T
The workshop covered text encoding, character encoding, regular expressions, search with regular expressions, search in unannotated corpora, simple text search, and search in annotated corpora with a corpus query language (for instance CQP) and search in XML documents using XQuery.

Course: Comparing Corpora
Peter Fankhauser (Institute for the German Language, Mannheim); Hannah Kermes (Saarland University, Saarbrücken); Elke Teich (Saarland University, Saarbrücken)
July 28 – August 1, 2014
European Summer School in Digital Humanities, "Culture & Technology"
http://www.culingtec.uni-leipzig.de/ESU_C_T
The course covered different aspects of corpus analysis regarding the comparision of corpora/subcorpora in the area of register analysis, contrastive analysis (mostly German English) including choosing the right corpora and features, specific aspects of corpus query, post-processing of extracted linguistic evidence, statistical analysis, classification and visualization.

Workshop: Spoken Language
Lecturers: Chr. Draxler (LMU Munich, Germany), Timm Lehmberg (Universität Hamburg, Germany)
The European Summer University in Digital Humanities
July 28 – August 1, 2014, Leipzig

URL: http://www.culingtec.uni-leipzig.de/ESU_C_T/node/371
Spoken language is fundamental to human communication. Research and technology development in spoken language require high quality speech recordings, fine-grained time-aligned annotations, and long-term and controlled access to the data.
This workshop covered the both the technological and research aspects of the creation and exploitation of speech resources.

1. In the technology part of the workshop (week 1), audio and video recording equipment, data and metadata formats, and the practical, social, and legal aspects of speech recordings were presented.

2. In the exploitation part of the workshop (week 2), techniques for basic annotations as well as domain-specific annotation schemes will be discussed; furthermore, an in-depth introduction to analysis methods and tools were given.

Workshop: Legal Aspects of Research Data Collection and Sharing
Paweł Kamocki (Institute for the German Language, Mannheim); Peter Lampen (ISAS Dortmund); Andreas Witt (Institute for the German Language, Mannheim)
September 15, 2014
Arbeitskreis Forschungsdaten der Leibniz-Gemeinschaft
Berlin, Germany

Workshop: CLARIN Community Session: Standard Community of practice
Andreas Witt (Institute for the German Language, Mannheim), Karlheinz Mörth (Austrian Academy of Sciences, Vienna), Maik Stührenberg (Institute for the German Language, Mannheim)
DARIAH General VCC meeting
September 18, 2014, Rome, Italy
`https://dariah.eu/activities/general-vcc-meetings/4th-general-vcc-meeting/programme/community-sessions.html`https://dariah.eu/activities/general-vcc-meetings/4th-general-vcc-meeting/programm sessions.html
In the proposed session the CSC presented its approaches to both providing information about of de-jure and de-facto standards and convincing the CLARIN centers to make use of designated standards for specific tasks. Moreover, the different standards used in the CLARIN community were discussed.

Workshop: Zweiter CLARIN-D Workshop zum Arbeitspaket 5: Dienste und Ressourcen
Volker Boehlke (Leipzig), Susanne Haaf (Berlin), Axel Herold (Berlin), Bryan Jurish (Berlin), Timm Lehmberg (Hamburg), Thorsten Trippel (Tübingen) , Kai Zimmer (Berlin)
2. DTA- & CLARIN-D-Konferenz und Workshop: Textkorpora in Infrastrukturen für die Geistes- und Sozialwissenschaften
November 17–18., 2014, Berlin (BBAW)
http://www.deutschestextarchiv.de/veranstaltungen/DTAClarinDConf2014
The workshop was a follow-up of the CLARIN-D/AP5-Workshop from January 2013. It presented new developments of CLARIN-D compatible language resources with respect to compilation and analysis.

Workshop: Introduction to metadata and archiving using the LAT tools
Alexander König (MPI Nijmegen)

Primera Escuela Documentaciön y Tipologïa Lingüística Alemania - Mexico
March 23-28 2015 Morelia, Mexico
Introduction course in archiving linguistic fieldwork data using the LAT tools, especially Arbil for creating and editing metadata and LAMUS for uploading the data and integrating it into the archive.

4.3 Tutorials and Courses

Tutorial: "CLARINifizierung" von Ressourcen – Anforderungen, Beispiele und Erfahrungen.
Volker Boehlke
DTA- & CLARIN-D-Konferenz und Workshop: Textkorpora in Infrastrukturen für die Geistes- und Sozialwissenschaften
November 17, 2014, Berlin
http://www.deutschestextarchiv.de/veranstaltungen/DTAClarinDConf2014# programm
This tutorial during the DTA- & CLARIN-D-Konferenz in Berlin addressed the topic of integrating new resources into the CLARIN-D infrastructure. After introducing important pillars of the CLARIN-D infrastructure, requirements for the "CLARINification" of data were described. In this context information necessary for the estimation of workload were given. Finally examples and experiences of successful integration processes were presented.

Course: Aufbau und Analyse historischer Korpora. Blockveranstaltung (Übung)
Susanne Haaf
February 2.–5., 2015, Universität Regensburg

Tutorial: Computerlinguistische Werkzeuge für die Sozialwissenschaften
André Blessing, Jonas Kuhn (IMS, University of Stuttgart)
Computerlinguistische Methoden der Inhaltsanalyse in den Sozialwissenschaften: Forschungspraktische Herausforderungen, Werkzeuge und Technologien (pre-conference workshop at the DHd 2015)
February 23 and 24, 2015, Graz (Austria)
https://dhd2015.uni-graz.at/de/ `http://www.uni-stuttgart.de/soz/ib/forschung/Forschungsprojekte/eIdentity.html`http://www.uni-stuttgart.de/soz/ib/forschung/Forschungsprojekte/eIdentity.html
The tutorial was part of the project workshop of the BMBF-funded project eIdentity (Multiple kollektive Identitäten in internationalen Debatten um Krieg und Frieden seit dem Ende des Kalten Krieges. Sprachtechnologische Werkzeuge und Methoden für die Analyse mehrsprachiger Textmengen in den Sozialwissenschaften). We presented the Complex Concept Builder. Participants could work with it via a demo website. We collected valuable feedback from users which will be used to further improve the tool.

Course: Text Technology
Jens Stegmann (IMS, University of Stuttgart)
Winter semester 2014/2015, Stuttgart
One aspect of this course was to introduce students to foundational XML technologies (documents, DTDs, XML Schema, XSLT/Xpath) used for processing and interchanging documents. The course also covered linguistic annotation, different types of pertinent resources, related standardization efforts and demonstrations of CLARIN-specific applications, e.g., WebLicht and the VLO.

Course: Arbeit mit Korpora gesprochener Sprache: Transkription, Annotation, Analyse

Hanna Hedeland, Daniel Jettka, Timm Lehmberg (HZSK)
Winter semester 2014/2015, University of Hamburg
http://www.slm.uni-hamburg.de/ifg1/Lehrplan-2/WiSe-2014-2015/KVV-WS1415-DSL.pdf
Introduction to approaches, methods and terms spoken language corpora and (transcription) standards in the context digital research infrastructures.

Course: Methoden und Praxis der Korpuslinguistik
Hanna Hedeland, Daniel Jettka, Timm Lehmberg (HZSK)
Summer Semester 2015, University of Hamburg
http://www.slm.uni-hamburg.de/ifg1/Lehrplan-2/SoSe-2015/Lehrplan-SoSe2015.pdf
Introduction to basic approaches, methods and terms of corpus linguistics with a focus on in the use digital research infrastructures.

4.4 Demos and Presentations

Presentation: Corpus Query Lingua Franca, Part I: Metamodel
Piotr Bański (Institute for the German Language, Mannheim); Elena Frick (Institute for the German Language, Mannheim); Andreas Witt (Institute for the German Language, Mannheim)
Juni 24, 2014
ISO / TC 37 and SCs meetings
Berlin, Germany

Presentation: Nutzung der Forschungsinfrastruktur CLARIN in den Geisteswissenschaften
Gerhard Heyer und Volker Boehlke
Workshop des ESF Verbundprojekts "Wissensrohstoff Texts""
Leipzig, 30. Juni 2014
The talk gave a short introduction to the notion of a research infrastructure in the humanities and how it can be used to select and process text ressources for historical languages.

Presentation: Should Miners go on Strike? Legal Issues in Text and Data Mining
Paweł Kamocki (Institute for the German Language, Mannheim)
33rd ATRIP Congress
July 8, 2014, Montpellier, France

Presentation: CLARIN: Resources, Tools, and services for Digital Humanities Research
Erhard Hinrichs (Eberhard Karls Universität Tübingen), Steven Krauwer (CLARIN ERIC)
Digital Humanities
July 9, 2014, Lausanne
http://dh2014.org/

Demonstration: WebLicht im Rahmen des Tages der Wissenschaft an der Universität Stuttgart
Kerstin Eckart, Anita Ramm, Jens Stegmann (IMS, University of Stuttgart)
July 12, 2014, Stuttgart
http://www.uni-stuttgart.de/tag/2013/`http://weblicht.sfs.uni-tuebingen.de/weblichtwiki/index.php/Main_Page`http://weblicht.sfs.uni-tuebingen.de/weblichtwiki/index.php/Main_Page
At the science day, research teams from the University of Stuttgart present demonstrations and en-

gage in discussions with interested people from outside of the University. As every year, the Stuttgart CLARIN-D team presented WebLicht with a focus on the Stuttgart web services.

Demo: Virtual Language Observatory
MPI-PL Nijmegen
CLARIN-D Kiosk at the European Summer University in Digital Humanities 2014
July 29 and 30 2014, Leipzig
http://www.culingtec.uni-leipzig.de/ESU_C_T/node/102
Demonstration of the Virtual Language Observatory in the CLARIN-Kiosk. Guidance for scientists to work with the observatory, and supply data for it.

Demo, Presentation: Adaptable Tools for Information Extraction and Visualization
Kerstin Eckart (IMS, University of Stuttgart)
CLARIN-D Kiosk at the European Summer University in Digital Humanities 2014
July 31 and August 1,2014, Leipzig
http://www.culingtec.uni-leipzig.de/ESU_C_T/
On two days of the ESU, we presented two tools developed at the CLARIN-D center Stuttgart: the Textual Emigration Analysis (TEA) and the Interactiv plattform for Corpus Analysis and Research tools (ICARUS).

Demo: CLARIN Booth at the ESSLLI Summer School
Dörte de Kok (Eberhard Karls Universität Tübingen)
ESSLLI Summer School 2014
August 11 – August 22, 2014, Tübingen
http://www.esslli2014.info/
The CLARIN-D center Tübingen was present with an information booth at the ESSLLI Summer School. At the booth, participants were able to receive information about the tools and services CLARIN-D offers, e.g. in the form of demonstrations of different tools.

Presentation: Online Processing of Speech Resources
Christoph Draxler (LMU Munich, Germany)
Batumi Summer School 2014
September 10, 2014, Batumi

Presentation: Non-consumptive Use. For the re-definition of exclusive rights to make room for Text and Data Mining
Paweł Kamocki (Institute for the German Language, Mannheim)
International Scholars Conference on Intellectual Property Law (SCIPLaw)
September 17, 2014, Vienne, Austria

Presentation: Applying the TEI SIG CMC Proposal to Wikipedia Discussion Corpora
Eliza Margaretha (Institute for the German Language, Mannheim); Harald Lüngen (Institute for the German Language, Mannheim)
Technical Meeting of the TEI CMC SIG, 2014 DARIAH Virtual Competence Centre Conference
September 18, 2014, Rome, Italy

Presentation: EXMARaLDA. A system of computer assisted transcription and annotation of spoken language (Partitur Editor, EXAKT, COMA)
Thomas Schmidt (Institute for the German Language, Mannheim); Timm Lehmberg (University of Hamburg)
CLARIN-DK, Workshop: Best practices for multilingual spoken data in linguistic corpora
September 18, 2014, LANCHART Centre, Copenhagen, Denmark
http://info.clarin.dk/kurser/best-practices/

Presentation: Language Island Data
Thomas Schmidt (Institute for the German Language, Mannheim)
CLARIN-DK, Workshop: Dealing with multilingual spoken data: Corpus-based approaches
September 19, 2014 LANCHART Centre, Copenhagen, Denmark
http://info.clarin.dk/kurser/best-practices/
The presentation covered the data processing procedure of language island data (Australia German), with regard to a) digitalisation, documentation and alignment, b) the long-time storage in the CLARIN infrastructure and c) the cross-linking of these language island data with other language island corpora within the same archive and/or other archives as well as with reference corpora (standard and dialect).

Presentation: Legal Interoperability: Problems and Solutions in CLARIN ERIC
Paweł Kamocki (Institute for the German Language, Mannheim)
Research Data Alliance, Interest Group on Legal Interoperability
September 22, 2014, Amsterdam, the Netherlands

Presentation: Protection of Data Privacy in EUDAT Workflows
Paweł Kamocki (Institute for the German Language, Mannheim)
3rd European Data Infrastructure Conference (EUDAT)
September 25, 2014, Amsterdam, the Netherlands

Presentation: Legal Issues in Text and Data Mining
Paweł Kamocki (Institute for the German Language, Mannheim)
EUDAT (European Data Infrastructure) Conference (session on Open Access)
September 25, 2014, Amsterdam, the Netherlands

Presentation: Annotieren, Klassifizieren – Fallstudien im Anwendungsbereich Internetbasierte Kommunikation
Harald Lüngen (Institute for the German Language, Mannheim); Michael Beißwenger (Technical University of Dortmund); Eliza Margaretha (Institute for the German Language, Mannheim); Christian Pölitz (Technical University of Dortmund)
September 25, 2014
In: *Variational Linguistics, KobRA-Workshop*
Mannheim, Germany

Poster: Mining Corpora of Computer-mediated Communication: Analysis of Linguistic Features in Wikipedia Talk Pages using Machine Learning Methods
Harald Lüngen (Institute for the German Language, Mannheim), Michael Beißwenger (Technical

University of Dortmund); Eliza Margaretha (Institute for the German Language, Mannheim); Christian Pölitz (Technical University of Dortmund)
October 7, 2014
In: *Workshop "Natural Language Processing for Computer-mediated Communication / Social Media", Konferenz zur Verarbeitung natürlicher Sprache (KONVENS 2014*)
Hildesheim, Germany

Presentation: BAS CLARIN WebServices
Thomas Kisler (LMU Munich, Germany)
KONVENS 2014
October 10, 2014, Hildesheim

Presentation: CMDI 1.2: Improvements in the CLARIN Component Metadata Infrastructure
Twan Goosen and Thomas Eckart
CLARIN Annual Conference 2014
24.10.2014, Soesterberg
This presentation depicted changes and improvements in the upcoming CMDI 1.2 specification. In addition several recommendations were given on how to use the new capabilities of this standard.

Presentation: Virtual Language Observatory 3.0: What's New?
Twan Goosen and Thomas Eckart
CLARIN Annual Conference 2014
25.10.2014, Soesterberg
This presentation illustrated the changes of the new VLO 3.0 release. Focus of the poster presentation was on the new interface and the improved metadata extraction process.

Presentation: Build your own Treebank
Daniël de Kok, Dörte de Kok, Marie Hinrichs (Eberhard Karls Universität Tübingen)
CLARIN Annual Conference 2014
October 25, 2014, Soesterberg
https://www.clarin.eu/event/2014/clarin-annual-conference-2014-soesterberg-netherlands

Demo: TeLeMaCo - A tool for the dissemination of teaching and learning materials
Hannah Kermes, Jörg Knappen, José Manuel Martínez Martínez, Elke Teich, Mihaela Vela (Universität des Saarlandes)
CLARIN Annual Conference 2014
October 25, 2014, Soesterberg
https://www.clarin.eu/event/2014/clarin-annual-conference-2014-soesterberg-netherlands
A life presentation of the TeLeMaCo service focussing on the ease of use.

Presentation: CLARIN's Virtual Language Observatory (VLO) under scrutiny: The VLO taskforce of the CLARIN-D centres
Susanne Haaf (BBAW), Peter Fankhauser (IDS), Thorsten Trippel (Eberhard Karls Universität Tübingen)
CLARIN Annual Conference 2014
October 25, 2014, Soesterberg
https://www.clarin.eu/event/2014/clarin-annual-conference-2014-soesterberg-netherlands

The presentation focused on the development of Virtual Language Observatory that has been designed as the central platform for primary access to the diverse resources and tools provided by CLARIN.

Presentation: Informatik und Digital Humanities: Chancen und Herausforderungen eines ungeklärten Verhältnisses
Erhard Hinrichs (Eberhard Karls Universität Tübingen)
Workshop "Informatik und die Digital Humanities"
November 3, 2014, Leipzig
http://informatik-dh-workshop2014.informatik.uni-leipzig.de/

Presentation: Zugang zu verteilten Kollektionen
Erhard Hinrichs (Eberhard Karls Universität Tübingen), Thorsten Trippel (Eberhard Karls Universität Tübingen), Oliver Schonefeld (IDS)
Research Data Alliance Deutschland
November 11, 2014, Potsdam
https://europe.rd-alliance.org/events/research-data-alliance-deutschland-treffen

Presentation: Mapping the Right Legal Issues
Paweł Kamocki (Institute for the German Language, Mannheim)
November 11, 2014, Rome, Italy
EUDAT WG on Data Access and Re-Use Policies (DARUP)

Presentation: Kontextualisierung von Sprachressourcen und -technologie in der geisteswissenschaftlichen Forschung
Christoph Draxler (LMU Munich, Germany), Elke Teich and Hannah Kermes (both Universität des Saarlandes)
CLARIN-D Konferenz AP5
November 18, 2014, Berlin

Presentation: Generative Bayes'sche Modelle zur explorativen Analyse eines historischen Zeitungskorpus
Peter Fankhauser (Institute for the German Language, Mannheim)
Textkorpora in Infrastrukturen für die Geistes- und Sozialwissenschaften, DTA- & CLARIN-D-Konferenz, BBAW
November, 18 2014, Berlin, Germany
http://www.deutschestextarchiv.de/veranstaltungen/DTAClarinDConf2014

Presentation: Concepts for Conceptualized Speech Visualization
Christoph Draxler (LMU Munich, Germany)
Volkswagenstiftung Workshop"Visuelle Linguistik"
November 19-21, 2014, Hannover

Presentation: Metadata-related Standards: TEI Header and CMDI
Andreas Witt (Institute for the German Language, Mannheim)
November 24, 2014, University of Warsaw, Warsaw, Poland
http://www.delab.uw.edu.pl/metadata-related-standards-tei-header-and-cmdi/

Presentation: Standardised Text Editing for the Humanities
Andreas Witt (Institute for the German Language, Mannheim)
December 16, 2014, University of Giessen
The International Graduate Centre for the Study of Culture
https://www.uni-giessen.de/faculties/gcsc/events/andreaswittkl

Presentation: Text Editing in Accord with the Text Encoding Initiative (TEI)
Andreas Witt (Institute for the German Language, Mannheim)
December 17, 2014
The Basics and Recent Developments
University of Giessen: The International Graduate Centre for the Study of Culture
http://www.uni-giessen.de/faculties/gcsc/events/andreaswittmasterclass

Presentation: BAS CLARIN WebServices
Thomas Kisler (LMU Munich, Germany)
AUVIS Workshop Fraunhofer Institute
January 22, 2015, Bonn

Demo: BAS CLARIN WebServices and Speech Corpora Repository
Thomas Kisler (LMU Munich, Germany)
Digital Humanities 2015
February 24-27, 2015, Graz

Demo: CLARIN Booth at the DHd 2015
Thorsten Trippel (Eberhard Karls Universität Tübingen)
Digital Humanities deutschsprachiger Raum
February 23 – February 27, 2015, Graz
http://dhd2015.uni-graz.at/
Digital Humanities in the German speaking area is a conference of the humanities scholars interested in using digital solutions. Participants are potential CLARIN users already in testing phases, as they are experienced with services in testing state and can provide valuable feedback while applying the infrastructure. The conference booth was used for special sessions, hands on training, answering questions and establishing contact.

Presentation: WebLicht: Bombardieren bevor die Services explodieren
Daniël de Kok, Wei Qiu, Marie Hinrichs (Eberhard Karls Universität Tübingen)
Digital Humanities deutschsprachiger Raum
February 25, 2015, Graz
http://dhd2015.uni-graz.at/

Presentation: Digitale Geisteswissenschaften und Informatik – Modelle der Zusammenarbeit
Erhard Hinrichs (Eberhard Karls Universität Tübingen)
Digital Humanities deutschsprachiger Raum
February 25, 2015, Graz
http://dhd2015.uni-graz.at/

4.5 Other

Special Session: Research Infrastructures and Resources for Processing Natural Language (CLARIN-D)
Erhard Hinrichs (Eberhard Karls Universität Tübingen), Thomas Kisler (BAS), Thorsten Trippel (Eberhard Karls Universität Tübingen)
Konvens
October 10, 2014
http://www.uni-hildesheim.de/konvens2014
At the KONVENS conference targeting an audience of computational linguists and language technologists, CLARIN presented the infrastructure and contributions for this user group. The group is especially relevant as they can contribute tools and resources that could be reusable by the CLARIN community.

Meeting: Linguistics and Computer Mediated Communication (CMC) SIG
Piotr Bański (Institute for the German Language, Mannheim); Andreas Witt (Institute for the German Language, Mannheim)
October 23, 2014
The fifth meeting of the Special Interest Group "TEI for Linguists", the 2014 Annual Meeting of the TEI Consortium
Chicago, Illinois, USA
http://tei.northwestern.edu/sig-meetings/

5 TeLeMaCo

TeLeMaCo is a collaborative portal for training and teaching materials relevant in linguistics and digital humanities hosted at the CLARIN-D center at Saarland University in Saarbrücken. The portal is easy to use both for casual users who search for teaching and training material and for community members who want to contribute descriptions of their materials. We collect structured metadata of the described resources to provide advanced search and to integrate them in the wider CLARIN framework of resources.

For language resources and tools, there is a growing number of repositories, and there are platforms to search the metadata of many collections at once, like the Virtual Language Observatory (VLO) [Van Uytvanck et al.(2012)]. There are also web services and chaining tools like WebLicht [Hinrichs et al.(2010)] that provide facilities for text and speech processing and support processing pipelines for a wide variety of scientific tasks.

On the other hand, the documentation and teaching materials remain scattered over many places, including institutional web pages, YouTube channels, or software repositories like sourceforge or Bitbucket. A common interface to access and search those teaching and learning materials was lacking when we started the design of our service.

We developed the Teaching and Learning Material Collection (TeLeMaCo) to overcome this situation. Our approach is community driven as we collect descriptions of relevant materials contributed from all over the world in our service.

More details about the content and the structure of TeLeMaCo can be found in the previous reports R8.2, R8.3.

5.1 Additional benefits

The contents of TeLeMaCo are crawled and indexed by the big search engines (Google, MSN, Yahoo, Yandex, Baidu). This has two effects to materials added to TeLeMaCo: Some users will find the TeLeMaCo display page in the search engine of their choice and go on to the wanted material, and the page rank of the original page is boosted (leading users directly from the search engine to the material). The display page of a teaching or learning material has internal weblinks for the authors and keywords weaving a web of related resources. This allows the user to navigate to other materials for the same tool, the same task, or by the same author.

5.2 Evaluation

After a phase of internal testing within the CLARIN-D project, TeLeMaCo went public in September 2013 and was announced at GSCL 2013 [Amoia et al.(2013)]. Since then, a steady trickle of descriptions was added to TeLeMaCo, now holding a total of 134 materials. Most of the contributions still come from members of the CLARIN-D project, but we start seeing submissions from other places, too.

We see in the logs that users from all over the world start consulting TeLeMaCo for teaching and learning materials. It was a surprise for us to see calls to a specific description directly without prior searching or browsing. These hits are coming from users being directed to TeLeMaCo by a search engine.

Since September 2014 TeLeMaCo is listed in the large directory at LINSE (Linguistik-Server Essen). LINSE is a German language portal to all kinds of materials and services around linguistics.

5.3 Comparison with other services

We are aware of two collections of teaching resources for the documentation of endangered languages. The E-MELD School of Best Practices [1] is a project supported by the LINGUIST list. Resources are added by the project team, and although there was little activity since 2007 the project is still alive. There are short descriptions of the materials in free text format and there is a search function.

The Resource Network for Language Documentation (RNLD) hosts a more up-to-date list of resources for language documentation. Materials are described in free text format. They provide a full text search over the whole website.

The project DARIAH-DE has launched a service called Schulungsmaterial-Sammlung in July 2014. The target group are Digital Humanities. The interface to this service is in German, materials are added by the staff members only. The materials have short textual descriptions and come with the following annotations: institution, media, title, tools, didactic type, discipline, language, date, keywords (up to three) chosen from the closed TaDiRAH [Perkins et al.(2014)] vocabulary, and license. There is a full text search over all the fields available.

5.4 Conclusion

We think that TeLeMaCo fills a gap in the existing ecosystem of language infrastructures by providing easy and quick access to teaching and learning materials. Descriptions of materials can be provided by everyone after registration at the service, avoiding a bottleneck in extending the service. TeLeMaCo provides structured metadata of the resources that can be further integrated in the CLARIN infrastructure. Both tools and available documentations benefit from being added to TeLeMaCo. They are not only findable through TeLeMaCo itself, but also their visibility in search engines is improved.

[1] http://emeld.org/school/index.html

Bibliography

[Amoia et al.(2013)] Amoia, M., Kermes, H., Knappen, J., Martínez Martínez, J. M., Teich, E., & Vela, M. 2013. Telemaco—a collaborative repository for training and teaching materials in linguistics. In *Proceedings of the International Conference of the German Society for Computational Linguistics and Language Technology (GSCL 2013)*, Darmstadt, Germany.

[Hinrichs et al.(2010)] Hinrichs, M., Zastrow, T., & Hinrichs, E. 2010. Weblicht: Web-based LRT services in a distributed escience infrastructure. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*, Valetta, Malta.

[Perkins et al.(2014)] Perkins, J., Dombrowski, Q., Borek, L., & Schöch, C. 2014. Building bridges to the future of a distributed network: From DiRT categories to TaDiRAH, a methods taxonomy for digital humanities. In *International Conference on Dublin Core and Metadata Applications 2014*, pages 181–183.

[Van Uytvanck et al.(2012)] Van Uytvanck, D., Stehouwer, H., & Lampen, L. 2012. Semantic metadata mapping in practice: the virtual language observatory. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC 2012)*, Istanbul, Turkey.