CLARIN-D Report R 8.2
Documentation of AP8 preparation activities

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

May 2014
## Contents

1. **Introduction**  
   
2. **Teaching and Learning Material Collection (TeLeMaCo)**  
   
3. **Training modules and tutorials**  
   
4. **Summer school**  

5. **Appendix**
1 Introduction

Within CLARIN-D the workpackage training and education plays a central role when it comes to dissemination and user orientation. It is essential for the success of the whole project to make potential users aware of the possibilities of the CLARIN-D services and tools for their research, and to make them familiar with the usage of these services and tools. In order to achieve this goal the members of the CLARIN-D consortium have continued to work on the three different contexts that we have identified for the workpackage training and education:

- integration of language resources and language tools into university degree programs,
- training of teaching staff
- qualification of researchers

During the third year of the project a particular emphasis was on the preparation of the summer school, which will take place from July 22 to August 1st. The summer school will be organized as a joint event with the well established European summer school in Digital Humanities “Culture & Technology” in Leipzig. A detailed report on these efforts can be found in section 4.

The centers continued to hold a variety of different smaller but not less important training modules and tutorials in various contexts, e.g.:

- a second event in the newly established series of CLARIN-D PhD-days at the Bavarian Archive for Speech Signal (Ludwig-Maximilian University in Munich) on speech databases
- several courses at different summer or winter schools (e.g. ESU DH C & T, INNET Summer School)
- tutorials in the context of larger conferences (e.g. the annual convention of the DGfS, the conference on Language Resources and Evaluation)
- individual workshops on specific topics (STTS-tagset, TEI-XML, historical corpora)
- courses in existing university programs

A report on the second CLARIN-D PhD days as well as an extensive list of tutorials, workshops and courses, demos and presentations is given in section 3 of the report.

We also continued work on the Teaching and Learning Material Collection (TeLeMaCo), a collaborative portal for training and teaching material, which was set up during the second year of the project, which has been continuously filled since then and is now publicly accessible.
2 Teaching and Learning Material Collection (TeLeMaCo)

The Teaching and Learning Materials Collection (TeLeMaCo) is a collaborative portal for training and teaching material relevant to linguistics and digital humanities. The material is provided with metadata including standard categories (such as author, title, publication year), as well as categories capturing pedagogically relevant features (target audience, objectives, prerequisites), technical features (language, media type), etc. TeLeMaCo is indexed by search engines (e.g., Google and Bing) and provides its own browsing and searching interfaces.

A more detailed description of TeLeMaCo can be found in the second-year project report of workpackage 8.

TeLeMaCo was completed and extensively tested in the first half of 2013. In the beginning it was open for CLARIN-D members only. In September 2013 it went public.

http://fedora.clarin-d.uni-saarland.de/hub/
TeLeMaCo has been continuously filled with training and teaching material covering Quick Starts, Handbooks, Dictionaries, Screencasts and Teaching Videos, Course Units and full Courses. TeLeMaCo is meant as a collaborative portal, thus, contributions do not only come from members of CLARIN-D but also from external users.

By now it contains more than 100 different material. Many of the material used in the training courses and tutorials given by the members of the CLARIN-D consortium are indexed and made available in this collection. The courses given by CLARIN-D members at the summer school will also be made available here.
3 Training modules and tutorials

3.1 CLARIN-D Doktorandentage (PhD days)

The CLARIN-D PhD days are intended as a small series of training events. The target group are young researchers (mostly PhD students). Each of the single events has a specific thematic focus (e.g. corpora; speech; historical data).

3.1.1 Second CLARIN-D PhD days: *Speech databases – from recording to publication*

**Date:** March 31, 2014  
**Venue:** IBZ Internationales Begegnungszentrum der Wissenschaft  
**CLARIN-D Team:** Fabian Bross, Christoph Draxler, Bernhard Jackl, Thomas Kisler, Uwe Reichel, Florian Schiel

The Bavarian Archive for Speech Signal, Institute of Phonetics and Speech organized the second event in this series which took place on March 31st at the Ludwig-Maximilian University in Munich. The title of the one-day workshop was *Speech data – from recording to publication*.

The topic of the workshop was the creation of speech databases targeted at advanced students, PhD students and PostDocs. 29 participants from all over Germany, the Austrian Academy of Sciences, and Zurich University attended the one-day event in Munich.

The workshop gave participants a hands-on experience in speech recording, annotation, metadata creation and finally storing the speech database in an online repository.

In the first session, given by Christoph Draxler, the participants were asked to perform speech recordings using the software Speech Recorder. These recordings were then transferred to a server and annotated.

Thomas Kisler presented WebMAUS, a web service for the automatic segmentation and labeling of speech. The speech recordings and the associated orthographic transcriptions are sent to the WebMAUS server, which then returns a high quality phonetic segmentation.

After lunch, Bernhard Jackl demonstrated his script COALA which takes as input simple text tables and produces CMDI-compatible metadata. Yet another script, presented by Uwe Reichel, ingests the signal files, annotations and metadata into a CLARIN-D compliant repository. The repository is harvested regularly, so that the new data becomes visible in the virtual language observatory and may be searched using the CLARIN-D federated content search.

The workshop ended with a discussion of the participants own research projects. This discussion was quite lively and productive. The BAS received important usability and feature feedback from the participants, while the workshop participants got to know the tools provided by BAS and CLARIN-D and how they could be used in their own work.
3.2 Tutorials and Courses

Tutorial: Curation Project 1 of the CLARIN-D Discipline-specific Working Group 1 “German Philology”
CLARIN-D Center BBAW and Deutsches Textarchiv
February 20, 2013, BBAW
Number of Participants: 15
On February 20th, 2013, a tutorial was held at the BBAW for participants working for the Curation Project 1 of the CLARIN-D Discipline-specific WG 1 “German Philology” and staff members of the DFG-funded project “Literatureexplosion und Intertextualität. Bedingungen und Merkmale der ‘Verschriftlichung des Lebens’ in Nürnberg” where practical questions and issues were addressed which had become apparent in the participants work following the DTA workflow and guidelines. The tutorial was a daylong event conducted by staff members of the DTA and CLARIN-D at the BBAW.

Course tutorials: MPI Internal LAT Training
Alexander König, Lecturer and organizer
May 13 and 15, 2013
MPI for Psycholinguistics, Nijmegen, The Netherlands

Tutorial: ELAN training for Movement Analysts and Dance Therapists
Han Sloetjes, Lecturer and organizer
May 17, 2013
The Language Archive, MPI for Psycholinguistics, Nijmegen, The Netherlands

Tutorial on ISOcat
Menzo Windhouwer, Lecturer
June 20, 2013
Utrecht, The Netherlands
TLA and CLARIN-NL
http://www.clarin.nl/node/1794
In this tutorial ISOcat and its use in the CLARIN infrastructure is introduced. Practical hands-on sessions were intermingled with presentations of best practices and guidelines.

Tutorial: Hands-on ELAN training
Han Sloetjes, Lecturer
INNET Summer School, Technological Approaches to the Documentation of Lesser-Used Languages
September 1–14, 2013
University Gniezno, Poland
http://www.innet-project.eu/events/summer-school-2013

Course tutorials: Metadata and Unicode/fonts
Alexander König, Lecturer
INNET Summer School on Technological Approaches to the Documentation of Lesser Used Languages
September 1–14, 2013
University Gniezno, Poland
http://www.innet-project.eu/events/summer-school-2013

Course tutorials: Archiving, IMDI metadata, ARBIL, LAMUS, AMS and LEXUS
Alexander König, Lecturer and organizer at Dobes Training Course
October 22–25, 2013
MPI for Psycholinguistics, Nijmegen, The Netherlands

Tutorial: ELAN training for Conversation Analysis
Han Sloetjes, Lecturer and organizer
November 13–15, 2013
Barcelona, Spain
Three-day course on discourse transcription with ELAN, conducted at the University of Barcelona. Participants were researchers working with corpora in Catalan, Spanish and English.

Annotation: Anwenderbedarf und Support Hamburg Centre for Language Corpora (HZSK)
November 28+29, 2013
On November 28th and 29th 2013, a workshop titled “Annotation: Anwenderbedarf und Support” (“Annotation: User demands and Support”) was conducted at the Hamburg Centre for Language Corpora (HZSK). The aim of the workshop was to bring together developers and users (primarily linguists, who formerly had requested help from the CLARIN-D help desk) of annotation and analysis software, who mutually exchanged their system demos to subsequently discuss requirements and possible solutions. The idea was to make the potential of the CLARIN-D infrastructure and help desk more known to these users and at the same time gain more information about their demands to improve the CLARIN-D help desk as well as the CLARIN-D tools and services.

EXMARaLDA practical training courses
Hamburg Centre for Language Corpora (HZSK)
January 11, 2014 (Warsaw)
February 27, March 5, March 13–15 2014 (Hamburg)
The training courses introduce EXMARaLDA (“Extensible Markup Language for Discourse Annotation”), a system of concepts, data formats, and tools for the computer assisted transcription and annotation of spoken language, and for the construction and analysis of spoken language corpora. For more information on EXMARaLDA training courses see: http://www.exmaralda.org/training.html

Course tutorials: MPI Internal LAT Training
Alexander König, Lecturer and organizer
January 30, 2014
MPI for Psycholinguistics, Nijmegen, The Netherlands

Tutorial of the “Sektion Computerlinguistik” (DGfS): Syntactic corpus annotation with We-
The target group of the tutorial are interested linguists without computer-linguistic background. Participants will learn how to use corpus linguistic methods and tools, e.g. the annotation and query of text corpora with different language resources and tools (e.g., WebLicht, TuNDRA, CQP), as well as statistical analysis visualization and interpretation of results. Participants learn:

- how to set up a pipeline for the syntactic annotation of texts;
- how to query text corpora for linguistic phenomena on different levels (word level, syntactic level)
- about possibilities for a (statistical) analysis and visualization of the extracted data and how the results of such a research may be interpreted

Tutorial “online speech and language resources”
Christoph Draxler (BAS), Daniel Jettka (HZSK), Thomas Kisler (BAS), Thomas Eckart (UL), organized by Bavarian Archive for Speech Signals
May 26–31, 2014

At the LREC 2014, CLARIN-D will present a half-day tutorial on the following topics:

Resource and Tool Creation
- Metadata: specification, creation and use of metadata
- Webservices: architecture and guidelines for webservice, case study of an existing web service

Online access to Resources
- Persistent Identifiers: motivation, concepts and implementation
- Repositories: architecture, harvesting, browsing, case study of an existing repository

The tutorial presentations will give an in-depth introduction into the topic and present show cases and sample applications from ongoing CLARIN-D activities.

Workshop: Query in Text Corpora
Andreas Witt (Institut für Deutsche Sprache, Mannheim)
July 23–August 2, 2013
European Summer School in Digital Humanities, “Culture & Technology”
http://www.culingtec.uni-leipzig.de/ESU_C_T

The workshop covers 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.
Susanne Haaf (Deutsches Textarchiv/CLARIN-D) gave an Introduction to TEI-XML at the International Summer School “Digitization and its Impact on Society”, held from September 29th until October 5th, 2013 at the TU Dresden. The workshop course encompassed an introduction on principles and possibilities of the TEI/XML annotation format.

Max-Planck-Institut für Physik komplexer Systeme Dresden,
Montag, 30. September 2013, 9:00–10:30 Uhr.

Workshop on ISOcat
Menzo Windhouwer, Lecturer
October 2, 2013
Utrecht, The Netherlands
http://www.clarin.nl/node/1899
TLA and CLARIN-NL
In this follow-up of the June ISOcat tutorial participants could share their experiences and ask questions. The guidelines on the use of ISOcat were presented again.

Workshop Corpus Annotation and Corpus Query with WebLicht and CQPweb
Hannah Kermes (Universität des Saarlandes)
October 8, 2013
Regensburg
The goal of the workshop was to give an overview of the functionalities and possibilities of WebLicht and CQPweb and to provide hands-on experience in their usage.

Workshop INNET 2nd regional archives workshop
Paul Trilsbeek and Alexander König, Lecturer and organizer
October 28–30 2013
MPI for Psycholinguistics, Nijmegen, The Netherlands
The INNET regional archivists workshop aimed at informing archivists of the various regional archives that make use of the LAT archiving framework about the latest developments within CLARIN and other research infrastructure projects, as well as about new developments in the LAT software itself.

Workshop: Das STTS-Tagset für Wortartentagging: Stand und Perspektiven – 3. Workshop
Heike Zinsmeister (IMS, Universität Stuttgart)
December 9, 2013, Hildesheim
http://www.uni-hildesheim.de/fb3/institute/iwist/
The third of three workshops jointly organized by the CLARIN-D centers SfS Tübingen and IMS Stuttgart in cooperation with IwiST Hildesheim. The focus of the workshops is on the documentation and availability of STTS “dialects”, as well as planning further STTS modifications and developments.

**Workshop: NEUROGES – ELAN gesture coding workshop**
Han Sloetjes, Lecturer and organizer
December 12–13, 2013
Fribourg, Switzerland
[http://www.unifr.ch/psychotherapie/neuroges](http://www.unifr.ch/psychotherapie/neuroges)
Seminar two (of three) of NEUROGES: a coding system for the empirical analysis of hand movement behavior and gesture as a reflection of cognitive, emotional, and interactive processes.

**Course: Language Documentation, Linguistic Annotation and Archive Building**
Sebastian Drude, Lecturer and organizer of this course at LOT Winterschool
January 14–16, 2013
Leiden University
In the last ten years or so Language Documentation has emerged and quickly developed, partly as a response to the imminent risk of losing the major part of the world’s linguistic diversity within a few generations. This new field of linguistics is in the intersection of field research and language technology, as the general goal is to create broad, lasting, multimodal and multipurpose digital corpora of language use, with a focus on natural speech in its cultural context. This course gives an introduction to language documentation, its background and general methodology, and focuses on some key topics, such as relationships with speakers, linguistic annotation of audio and video recordings, and several methodological and technical issues in the building of digital archives.

**Lecture: The Language Archive at the MPI Nijmegen and the future of Endangered Language at Digital Data**
Sebastian Drude, Lecturer at Atuais Tendências e Potenciais na Documentação Linguística (Workshop und Sommerschule zu Sprachdokumentation)
April 2–5, 2013
Landes-Universität Campinas (Unicamp), Brasilien
[http://www.iel.unicamp.br/destaques/arquivos/CartazATPDL.pdf](http://www.iel.unicamp.br/destaques/arquivos/CartazATPDL.pdf)
Natural language is the medium expressing collective and personal knowledge of any social group. We know, however, that the cultural and linguistic diversity found throughout the world will be drastically reduced in the coming decades. In order to preserve this valuable data on human knowledge, natural language and its uses, and ultimately on human language faculty, immediate and collective afford is required. It is at this point that Language Documentation as research field comes into play.

**BS Grundlagen und Methoden der Sprachdokumentation (Foundations and methods of language documentation)**
Sebastian Drude, Lecturer
July 22–August 2, 2013
Goethe-Universität, Frankfurt am Main
Course lecture Metadata management
Course tutorial Metadata
Paul Trilsbeek, Lecturer
INNET Summer School on Technological Approaches to the Documentation of Lesser Used Languages
September 1–14 2013
University Gniezno, Poland
http://innet-project.eu/events/summer-school-2013
The INNET summer school was targeted at young researchers and focused on standards, state-of-the-art methodologies and advanced tools for the digital documentation and archiving of languages.

Lecture: The Language Archive as a center of the CLARIN infrastructure
Sebastian Drude, Lecturer
at The second INNET conference on digital language archiving (Best practices in digital language archiving of language and music data)
September 6–7, 2013
University Gniezno, Poland

Courses: Shoebox (Toolbox) introduction & practice
Sebastian Drude, Lecturer at Dobes Training Course
October 22–25, 2013
MPI for Psycholinguistics, Nijmegen, The Netherlands

Courses:

- Introduction to archiving,
- Audio recording, Digital audio, handling, Tools and practice,
- Video recording, Digital video handling,
- Video tools and practice,
- Power Management, A/V file formats, A/V compression,

Paul Trilsbeek, Lecturer and organizer at DOBES training course
October 22–25, 2013
MPI for Psycholinguistics, Nijmegen, The Netherlands
The DOBES training course focuses on the technical aspects of language documentation. The topics covered in this training course were Audio Recording, Video Recording, Digital Video Handling, working with the ELAN annotation tool, working with Toolbox, making metadata descriptions, LAMUS archive upload and management system, AMS access management system, Audio and Video compression, Power and Storage Management in the Field, Unicode/Keyman, LEXUS lexicon tool and Workflow Principles.
Seminar/Course: Digitale Edition von Funeralschriften des 17./18. Jahrhunderts aus dem Be-
stand der Staatsbibliothek zu Berlin (Digital Edition of Funeral Sermons from the 17th/18th
Century from the holdings of the Staatsbibliothek zu Berlin)
CLARIN-D Center BBAW and Deutsches Textarchiv
Winter Term 2013/14, Freie Universität Berlin
http://www.fu-berlin.de/vv/de/lv/123794?m=189226&pc=74927&sm=69108

During the winter term of 2013/14 Susanne Haaf (Deutsches Textarchiv/CLARIN-D) and Matthias
Schulz (Deutsches Textarchiv) taught a course on the “Digital Edition of Funeral Sermons from the
17th/18th Century” in the context of the Interdisciplinary Master’s Program Editionswissenschaft at
the Faculty for German and Dutch Philology of the Freie Universität Berlin. Its aim was to impart
knowledge on methods of digital scholarly editing, ranging from the installment of the edition’s ra-
tionale with regard to the specifics of the source material and the anticipated expectation of its users

Course: Text Technology
Jens Stegmann (IMS, Universität Stuttgart)
Winter Term 2013/2014, Stuttgart
http://tinyurl.com/osapsw9

One main aspect is to introduce students to foundational XML-related technologies (documents,
DTDs, XML Schema, XSLT/Xpath) beneficial for processing of and the interchange of textual doc-
uments. The course also covers issues in linguistic annotation, different types of pertinent resources,
related standardization efforts and demonstrations of CLARIN-specific applications, e.g. WebLicht

CLARIN-D Resources used in courses
WebLicht and Tundra used in teaching

• Prof. Dr. Angelika Storrer at the University of Mannheim uses WebLicht in her course “Digitale
Sprachressourcen in der Linguistik”.

• Prof. Dr. Gottfried Zimmermann at the Stuttgart Media University used WebLicht and Tundra
in his course “User Interaction”

http://www.fu-berlin.de/vv/de/lv/123794?m=189226&pc=74927&sm=69108
5 http://www.deutschestextarchiv.de/
6 The results of the previous course from winter term 2012/2013 are already published on the web-
based quality assurance platform DTAQ: http://www.deutschestextarchiv.de/DTAQ/book/show/
lange_predigt_1643 http://www.deutschestextarchiv.de/DTAQ/book/show/hausen_predigt_1710
show/schele_predigt_1670
3.3 Demos and Presentations

Demonstration: WebLicht im Rahmen des Tages der Wissenschaft an der Universität Stuttgart
Andre Blessing, Kerstin Eckart, Jens Stegmann (IMS, Universität Stuttgart)
June 22, 2013, Stuttgart
[http://www.uni-stuttgart.de/tag/2013/](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)

On occasion of the science day, research teams from the University of Stuttgart present demonstrations and engage in discussions with interested people from outside of the University. Hands-on approaches - excitement, fun and eye-openers included! As every year, the Stuttgart CLARIN-D team presented WebLicht with a focus on the Stuttgart web services included in it.

Demonstration: WebLicht im Rahmen des Besuchs von Züricher Computerlinguisten am IMS
Kerstin Eckart (IMS, Universität Stuttgart)
November 7, 2013, Stuttgart

A short demonstration of WebLicht with a focus on the Stuttgart web services provided.

Presentation: CLARIN in Stuttgart: Metadaten, eHumanities und trainierbare Werkzeuge.
Vortrag im Kolloquium der Abteilung Korpuslinguistik und Morphologie am Institut für deutsche Sprache und Linguistik, Humboldt-Universität zu Berlin.
Kerstin Eckart (IMS, Universität Stuttgart)
November 27, 2013, Berlin.

A presentation involving a demonstration of WebLicht, the VLO and the Stuttgart CLARIN repository. Focus on CMDI metadata for (re-)trainable tools and pertinent resources.

TeLeMaCo—a collaborative repository for training and teaching materials in linguistics
Jörg Knappen, Presenter
International Conference of the German Society for Computational Linguistics and Language Technology (GSCL 2013)
September 26, 2013
Technische Universität Darmstadt, Germany

Presentation: Textual Emigration Analysis
Annual CLARIN Meeting 2013 in Prague
Andre Blessing (IMS, Universität Stuttgart)
October 21, 2013, Prague
[https://www.clarin.eu/content/showcases-i#emigration](https://www.clarin.eu/content/showcases-i#emigration)

Historians, literary scientists, and others are interested in the semantic interpretation of text. With automatic pre-processing of texts, e.g. named entity recognition, coreference resolution, and dependency parsing, relevant semantic relations can be extracted. The Stuttgart tools extract relations of migration, i.e. name and job of the migrant and date, reason and destination of migration. A graphical
interface showing timeline and geographic migrations will be implemented.

**Presentation: IMS Web Services — Toolchains and Applications**  
Annual CLARIN Meeting 2013 in Prague  
Jens Stegmann (IMS, Universität Stuttgart)  
October 22, 2013, Prague  

IMS web services are fully integrated into the CLARIN infrastructure and, e.g., available via the WebLicht platform. Properties: modular and flexible vs. monolithic toolchains. Aspect of reusability in other application contexts.

**Presentation New Developments in Arbil**  
Peter Withers, Presenter  
Paradisec RRR Program (Research, records and responsibility, RRR: Ten years of the Pacific and Regional Archive for Digital Sources in Endangered Cultures, PARADISEC)  
December 2–4, 2013  
University of Melbourne, Australia  

**Presentation, New Developments in KinOath Kinship Archiver**  
**Presentation, New Developments in Arbil Metadata Tool**  
Peter Withers, Presenter  
Department of Linguistics, University of Hawaii  
January 14, 2014

University of Hawaii - Manoa Campus  
Peter Withers, KinOath Kinship Archiver is a kinship application designed to be extremely flexible and culturally non nonspecific. There have been a number of additions to the application in the last year such as easier selections on the diagram, directed relations and more flexible field creation. This talk will discuss how to use the application as well as the recently added features and some future features[http://www.ling.hawaii.edu/tuesday-seminar-series-spring-2014](http://www.ling.hawaii.edu/tuesday-seminar-series-spring-2014)

**Presentation, New Developments in KinOath Kinship Archiver**  
**Presentation, New Developments in Arbil Metadata Tool**  
Peter Withers, Presenter  
SILS Conference (The 21st Stabilizing Indigenous Languages Symposium)  
January 15–19, 2014  
University of Hawaii at Hilo  

**Presentation: ARBIL**  
**Presentation: YAMS, CMDI Metadata Search**  
Peter Withers, Presenter at CMDI tutorial CLARIN NL  
March 4, 2014  
Utrecht University, The Netherlands  
Presentation: CMDI & interoperability: ISOcat and RELcat; Semantic Interoperability
Menzo Windhouwer, Presenter
CMDI tutorial CLARIN NL
March 4, 2014
Utrecht University, The Netherlands
http://www.clarin.nl/node/1969

Presentation: VLO (Virtual Language Observatory): faceted browsing, keyword search
Alexander König, Presenter at CMDI tutorial CLARIN NL
March 4, 2014
Utrecht University, The Netherlands
http://www.clarin.nl/node/1969

Presentation: Introduction to CLARIN-D
Lecturer/Presenter (CLARIN-D Center): Gerhard Heyer, Leipzig
1st Workshop of the project eXChange - Exploring Concept Change and Transfer in Antiquity
March 20, 2013 in Leipzig
Presentation of the goals and services available in CLARIN to serve the scientific community of classical studies.

3.4 Other

Section “CLARIN-D: eine Forschungsinfrastruktur für die Geistes- und Sozialwissenschaften”
Christoph Draxler (BAS), Thomas Eckart (UL), Thomas Kisler (BAS), Jörg Knappen (UdS), Thorsten Trippel (UT), organized by Bavarian Archive for Speech Signals
March 25–28, 2014
Digital Humanities im deutschsprachigen Raum (http://www.dhd2014.uni-passau.de), Passau
At the DHd 2014, a section is a focused set of talks over 90 minutes. In the first contribution to this section, given by Thorsten Trippel, CLARIN-D was presented first in the context of research infrastructures. Then, in the following presentation by Thomas Eckart and Jörg Knappen, the steps necessary to integrate new resources into the CLARIN-D infrastructure were outlined, with a strong emphasis on the support provided by CLARIN tools and members. In the final presentation by Thomas Kisler, the automatic segmentation system WebMAUS was presented as a successful show case. Approx. 40 people attended the section, and the discussion was lively.
4 Summer school

ESU DH “Culture & Technology” and CLARIN-D will organize a joint Summer School in 2014 at the University of Leipzig (http://www.culingtec.uni-leipzig.de/ESU_C_T/). The summer school will take place from the 21st of July (arrival) to the 2nd of August (departure). Through this collaboration we seek to cover a broad spectrum of Digital Humanities topics. A special focus will be on the use of language technology in Digital Humanities. Organizing a joint summer school with ESU DH “Culture & Technology” has several advantages. We can profit from the experience in organizing such an event, and the fact that the summer school is already known by many students and researchers. Thus, we will be able to address a broader range of prospect users of the CLARIN-D infrastructure and services.

4.1 Concept and mission

The Summer School wants to provide a place not only for learning but also for a get-together of students, young scholars and researchers from a variety of different scientific areas (Arts and Humanities, Library Services, Engineering and Computer Sciences). The goal is to foster the interdisciplinary exchange of knowledge and experience between equal partners in a multilingual and multicultural context creating the basis for future cooperations on project-bases and network-building across disciplines, countries and cultures.

The Summer School aims to provide a stimulating environment for discussing, learning and acquiring knowledge and skills in methods and techniques connected to Humanities Computing. Possibilities and limitations, consequences and implications of the application of computational methods and tools are discussed and insights into the complexity of Humanities data and the challenges the Humanities present for computer science and engineering and their development are given.

The Summer School will offer Humanities students the possibility to gain practical knowledge of the application of computational methods to the digitization, description, analysis and production of Humanities contents and artifacts (languages, texts, images, etc.), to discuss related theoretical questions and to forge new perspectives on the study and preservation of languages, cultures and cultural memory and the translation between cultures.

Students of Computer and Engineering Sciences will be given the opportunity to acquire insights into the nature of Humanities data. They will get acquainted with those areas in Arts and Humanities in which computational methods are employed and will learn to recognize the different needs research in the Humanities poses to computational methods. They will be confronted with the challenges work with diffuse and extremely complex data presents for soft- and hardware solutions.

For CLARIN-D the Summer School is a platform to present its resources and tools to (young) researchers from different areas of the Humanities, to give hands-on experience in the use and usage of these resources and tools, and to discuss (specific) research questions and problems with the participants. Here, the close contact with the participants in workshops, project and poster presentations, the CLARIN-D Kiosk plays an important role and provides ample opportunities for scientific discussions and knowledge exchange.
4.2 Program components and structure

The Summer School spans over 11 whole days. The program covers workshops, public lectures, regular project presentations, a poster session and a panel discussion. The program structure, in general, is as follows:

- morning: workshops
- afternoon: project presentations/poster sessions (open to the public)
- evening: lectures, panel discussion (open to the public)

Lectures are open to the public, and deal with broader topics of general interest. Project presentations have two goals: (i) give participants the possibility to present their own projects even if they are in a beginning stage, (ii) present larger (funded) projects of more advanced researchers and research groups. This panel discussion will deal with the topic digital Humanities in pedagogy.

As a new element of the Summer School we will organize a “CLARIN-D Kiosk”, where throughout the whole Summer Schools CLARIN-D centers will present their tools and services, as well as projects that have been realized with the help of these resources and technologies (e.g. curation projects). The “CLARIN-D Kiosk” will also be a facility for interested researchers and students to ask and discuss possibilities but also limitations of using language resources and language technology in their research.

4.3 Workshops

The Summer School offers a range of workshops on important areas of Digital Humanities and Language Resources. All workshops run in parallel throughout 11 days. Each workshop consists of a total of 16 sessions or 32 teaching hours. The number of participants in each workshop is limited to 12.

The workshops will be structured in two equal blocks of 16 teaching hours each. Participants can either take two blocks of one workshop or two blocks from different workshops. Participants who wish to switch between different workshops after the first week will need to demonstrate in their application their ability to participate in the other workshop. For an overview of the workshops see the table in the appendix. It will not be possible to register for one block only.

Wherever possible, the participants should be able to test the learned within their research field. Thus, participants are asked to bring relevant data from their projects.

Material (slides, sample data, ...) for each workshop will be made available in a Moodle. The Moodle will be used as a teaching environment during the Summer School.

The workshop program includes topics from different areas of Digital Humanities:

1. Alex Bia (Universidad Miguel Hernández, Elche, Spain): *XML-TEI encoding, structuring and rendering*

2. Andreas Witt (Institut für Deutsche Sprache, Mannheim, Germany): *Query in Text Corpora*

3. Peter Fankhauser (Institut für Deutsche Sprache, Mannheim, Germany) / Hannah Kermes (Universität des Saarlandes, Germany), Elke Teich (Universität des Saarlandes, Germany): *Comparing Corpora*
4. Susanne Haaf (Berlin-Brandenburgische Akademie der Wissenschaften), Christian Thomas (Berlin-Brandenburgische Akademie der Wissenschaften): *Historical Text Corpora for the Humanities and Social Sciences. Digitization, Annotation, Quality Assurance and Analysis*

5. Monica Berti / Simona Stoyanova / Giuseppe Celano (Humboldt Chair, University of Leipzig, Germany): *Open Greek and Latin*

6. Matt Munson (University of Leipzig, Germany): *Advanced Topics in Humanities Programming with Python*

7. Jan Rybicki (Uniwersytet Jagielloński, Kraków, Poland) / Maciej Eder (Uniwersytet Pedagogiczny, Kraków, Poland): *Stylometry: Computer-Assisted Analysis of Literary Texts*

8. Johanna Green (University of Glasgow, Scotland): *Editing in the Digital Age: Historical Texts and Documents*

9. Matthias Lang / Dieta Svoboda (University of Tübingen, Germany): *Space - Time - Object: Digital methods in archaeology*

10. Christoph Draxler (LMU München, Germany), Timm Lehmberg (Universität Hamburg, Germany): *Spoken Language*

11. Laszlo Hunyadi (Debreceni Egyetem / University of Debrecen, Hungary): *Multimodal Corpora: How to build and how to understand them*

12. Lynne Siemens (University of Victoria, Canada) / Jennifer Guiliano (University of Maryland, College Park, USA): *Large Project Planning and Management*

13. Ray Siemens (University of Victoria, Canada): *DH for Department Chairs and Deans*
### 5 Appendix

#### 5.1 Summer school workshop overview

<table>
<thead>
<tr>
<th>Workshop</th>
<th>Week 1</th>
<th>Week 2</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>XML-TEI encoding, structuring and rendering</strong></td>
<td>Introduction to XML-TEI text encoding</td>
<td>XML-TEI document structuring and rendering</td>
</tr>
<tr>
<td><strong>Query in Text Corpora</strong></td>
<td>Basic aspects of work with digital texts and construction of web corpora</td>
<td>Issues concerning the exploration of language corpora</td>
</tr>
<tr>
<td><strong>Comparing Corpora</strong></td>
<td></td>
<td>Aspects of corpus analysis regarding the comparison of corpora/subcorpora</td>
</tr>
<tr>
<td><strong>Historical Text Corpora for the Humanities and Social Sciences</strong></td>
<td>Digitization, Annotation, Quality Assurance and Analysis</td>
<td></td>
</tr>
<tr>
<td><strong>Open Greek and Latin</strong></td>
<td>EpiDoc and CTS/CITE Architecture</td>
<td>Linguistic annotations</td>
</tr>
<tr>
<td><strong>Advanced Topics in Humanities Programming with Python</strong></td>
<td>Machine Learning</td>
<td>Automatic Semantic Information Extraction</td>
</tr>
<tr>
<td><strong>Stylometry: Computer-Assisted Analysis of Literary Texts</strong></td>
<td>From Nothing to Networks</td>
<td>From Networks to Nearest Shrunkened Centroids</td>
</tr>
<tr>
<td><strong>Editing in the Digital Age: Historical Texts and Documents</strong></td>
<td>Editorial theory of literary and non-literary works, from manuscript and printed text, to the digital page</td>
<td>Practical application of editorial theory covered in Week One to a range of historical texts and documents</td>
</tr>
<tr>
<td><strong>Space - Time - Object: Digital methods in archaeology</strong></td>
<td>Geographic Information Systems</td>
<td>2 – 3D</td>
</tr>
<tr>
<td><strong>Spoken Language</strong></td>
<td>Technological aspects of the creation and exploitation of speech resources</td>
<td>Exploitation of speech resources</td>
</tr>
<tr>
<td><strong>Multimodal Corpora: How to build and how to understand them</strong></td>
<td>Introduction to multimodality in communication</td>
<td>Understanding multimodality in communication</td>
</tr>
<tr>
<td><strong>Large Project Planning and Management</strong></td>
<td>Introduction to Project Management</td>
<td>Advanced topics in Project Management</td>
</tr>
<tr>
<td><strong>DH for Department Chairs and Deans</strong></td>
<td>Survey and discussion of pragmatic DH basics and chief administrative issues related to supporting DH and those who practice it at their institution; audit of courses</td>
<td>Consultation and targeted discussion with the instructors and others in the group; audit of courses</td>
</tr>
</tbody>
</table>