Deliverable: R2.4
Fourth Year Activity Report
Liaison

Responsible: Erhard Hinrichs

© All rights reserved by the University of Tübingen on behalf of CLARIN-D

Editor: Thorsten Trippel
1 Introduction .............................................................................................................................................. 4

2 Liaison activities within CLARIN .............................................................................................................. 8
   Liaison activities with CLARIN ERIC ........................................................................................................ 8
   Liaison activities with foreign CLARIN partners .................................................................................... 9

3 Liaison with other projects and initiatives ............................................................................................... 10
   DARIAH-DE: Joint Technical Advisory Board .................................................................................... 10
   EUDAT – European Data Infrastructure .............................................................................................. 10
   Research Data Alliance ......................................................................................................................... 11
   Standardization ...................................................................................................................................... 11
   Liaison activities with other initiatives ................................................................................................. 11

4 User Liaison .............................................................................................................................................. 13

5 Further perspectives ................................................................................................................................. 14
1 Introduction

CLARIN-D report R2.4 “Fourth Year Activity Report – Liaison” documents liaison activities during the fourth year of the CLARIN-D project. It presents an overview of cooperation within the CLARIN-D project as a whole, as well as the cooperative efforts of the individual CLARIN-D centres.

The CLARIN-D project¹ (Web and Centre-based Infrastructure for the Social Sciences and Humanities, HSS) is motivated by a vision of providing humanities scholars with single-sign-on access to diverse repositories of language and textual data and the tools to work with them. It is a collaborative effort involving institutions and researchers at nine academic centres:

- Bavarian Archive for Speech Signals (BAS), Ludwig Maximilians University of Munich
- Berlin Brandenburg Academy of Sciences and Humanities (BBAW)
- Institute for the German Language, Mannheim (IDS)
- Max Planck Institute for Psycholinguistics, Nijmegen (MPI)
- University of Tübingen, Department of General and Computational Linguistics
- University of Hamburg, Hamburg Centre for Speech Corpora (HZSK)
- University of Leipzig, Institute of Computer Science
- Saarland University, Department of English and Translation Studies
- University of Stuttgart, Institute for Natural Language Processing

CLARIN-D is building on the achievements of the European CLARIN (Common Language Resource and Technology Infrastructure) initiative as well as the German national project D-SPIN (Deutsche Sprachressourcen-Infrastruktur). CLARIN and D-SPIN developed a set of common standards and goals for humanities research infrastructures as well as plans for the sustainable provision of digital resources and their long-term archiving.

The projects D-SPIN (Germany) and CLARIN (EU) were preparatory to CLARIN-D. D-SPIN’s funding was provided by the German Federal Ministry of Education and Research (BMBF), while CLARIN was funded directly by the European Commission. Both were part of the European Seventh Framework Programme (FP7)² as well as the ESFRI Roadmap³ (European Strategy Forum on Research Infrastructures).

¹ http://www.clarin-d.de/
³ http://ec.europa.eu/research/infrastructures/index_en.cfm?pg=esfri
FP7 was a strategic plan to encourage research and innovation in the European Union and associated states. It follows the first six framework programmes, which operated from 1984 through 2006, and covered the period from 2007 through 2013, encompassing roughly 50 billion euros of funding activities. It was replaced by Horizon 2020 in 2014.

ESFRI is an on-going European planning framework to coordinate research activities across Europe in order to establish competitive, open-access, continent-wide infrastructures designed to better serve research communities. It promotes the development of cross-border research facilities and services covering broad areas of scientific and academic interest.

ESFRI research infrastructure projects are planned with three sequential phases: a preparatory phase, an implementation phase, and an exploitation phase. D-SPIN was a preparatory phase project, during which plans for digital humanities infrastructure were developed through technical research, prototypes, models for training researchers, and through clarification of the legal framework for the project. CLARIN-D is the implementation phase successor to D-SPIN. Funding during the implementation phase of the ESFRI roadmap comes entirely from participating European countries that directly fund national projects to fulfil the roadmap’s goals.

D-SPIN’s European preparatory phase partner (CLARIN) has been replaced by CLARIN ERIC, which is structured as an association of national member projects. This new organization is a European Research Infrastructure Consortium (ERIC)⁴, – a new category of legal entity established in European Community law as part of ESFRI in order to facilitate the establishment and operation of joint research infrastructures of Europe-wide interest.

CLARIN ERIC currently has fourteen members, thirteen of whom are national project organizations joining together researchers and institutions in their respective states and one is an intergovernmental body:

- Austria⁵
- Bulgaria
- Czech Republic⁶
- Denmark⁷
- Estonia⁸

---

⁴ http://ec.europa.eu/research/infrastructures/index_en.cfm?pg=eric
⁵ http://www.oeaw.ac.at/icltt/node/14
⁶ http://lindat.mff.cuni.cz/lindat/
⁷ https://clarin.dk/clarindk/forside.jsp
⁸ http://keeleressursid.ee/en/
• Germany
• Greece
• Lithuania
• Netherlands
• Norway
• Poland
• Portugal
• Sweden
• The Nederlandse Taalunie, a joint Dutch-Flemish intergovernmental organization responsible for promoting and maintaining standards for the Dutch language

Membership is open to European Union countries and EU-associated entities such as intergovernmental organizations as well as individual centres. Several other countries are in the process of joining CLARIN ERIC.

CLARIN-D and the other national CLARIN participants are encouraged to cooperate and coordinate their activities with other projects, at the national and local levels as well as internationally. On the national level, CLARIN-D has aligned some of its activities with DARIAH-DE (Digital Research Infrastructure for the Humanities), another project on the ESFRI roadmap with a structure similar to that of CLARIN, focused on digital research infrastructure for arts and humanities research. DARIAH-DE is funded by the German Federal Ministry of Education and Research (BMBF) and is the German national participant in the DARIAH ERIC.

CLARIN-D members are also involved in a number of BMBF-funded eHumanities projects, often bringing their experience with CLARIN-D and its infrastructure to these initiatives. Furthermore, as academic centres and universities, all CLARIN-D members are involved in multiple national and international projects, and they are encouraged to promote and use the CLARIN-D infrastructure whenever it can serve the scientific goals of researchers.

9 http://www.clarin-d.de
10 http://www.clarin.gr/
11 http://www.clarin.nl
12 Observer, http://clarin.b.uib.no/
14 http://sweclarin.se/
15 http://taalunie.org/wat-doet-taalunie/activiteiten/clarin
In order to facilitate interaction with users and other initiatives in various Digital Humanities disciplines, CLARIN-D installed a dedicated Liaison Coordinator, Thorsten Trippel. This report encompasses the liaison and cooperative activities of CLARIN-D centres and CLARIN-D as a whole with CLARIN ERIC, with other CLARIN ERIC members, with DARIAH-DE, with BMBF eHumanities projects, and with other national and international initiatives.
2 Liaison activities within CLARIN

This section summarizes the liaison activities of CLARIN-D centres with CLARIN ERIC and foreign CLARIN centres. CLARIN-D centres have a great amount of autonomy to initiate contact with CLARIN ERIC and work together with foreign CLARIN centres within the framework of this project.

Liaison activities with CLARIN ERIC

CLARIN-D partners participate at all levels of CLARIN ERIC. Erhard Hinrichs of the University of Tübingen is a member of the CLARIN ERIC Board of Directors (BoD) in his capacity as chair of the CLARIN ERIC National Coordinators’ Forum (NCF). Dieter Van Uytvanck, who was the technical coordinator of CLARIN-D, is now the CTO of CLARIN ERIC and is member of the board of directors by virtue of being the chair of the Standing Committee for CLARIN Technical Centres (SCCTC). Dr. Helge Kahler (BMBF) is President of the CLARIN ERIC.

Other CLARIN-D personnel serve in various roles within CLARIN ERIC, including:

- Axel Herold of the Berlin-Brandenburg Academy of Sciences and Humanities (BBAW) is a member of the CLARIN ERIC Standards Committee as well as National Content Manager for CLARIN-D.
- Erik Ketzan of the Institut für Deutsche Sprache (IDS) is chair of the CLARIN ERIC Legal Issues Committee.
- Dörte de Kok of the University of Tübingen serves as administrative assistant of the National Coordinators’ Forum (NCF).
- Heike Stadler (IDS) serves as Deputy National Content Manager.
- Andreas Witt (IDS) is the Chairman of the CLARIN ERIC Standards Committee.

CLARIN-D centres and personnel participated in the CLARIN ERIC conference held in Soesterberg, The Netherlands, October 23rd-25th 2014, with the following contributions:

- **WebAnno: a flexible, web-based annotation tool for CLARIN**, a CLARIN-D curation project of the discipline-specific working group 7 “Applied Linguistics, Computational Linguistics”, with contributions by Richard Eckart De Castilho, Chris Biemann, Iryna Gurevych and Seid Muhie Yimam
- **The CLARIN-D Help Desk**, a CLARIN-D work package of the CLARIN-D centre in Hamburg, presented by Timm Lehmbrek
- **CMDI 1.2: Improvements in the CLARIN Component Metadata Infrastructure**, an international cooperation with contributions from Axel Herold (BBAW), Thomas Eckart (University of Leipzig), Oliver Schonefeld (IDS) together with the international

colleagues Twan Goosen, Menzo Windhouwer, Oddrun Ohren, Matej Ďurčo

- **Build your own Treebank**, a contribution by the CLARIN-D centre in Tübingen by Daniël de Kok, Dörte de Kok and Marie Hinrichs

- **Using Data Mining and the CLARIN Infrastructure to Extend Corpus-based Linguistic Research**, a contribution by users of the CLARIN-D infrastructure at the University of Dortmund, presented by Christian Pölitz and Thomas Bartz

- **Auxiliary Fronting in German Subordinate Clauses: A Synchronic and Diachronic Corpus Study Using the CLARIN Research Infrastructure**, an application of the CLARIN-D infrastructure, presented by Erhard Hinrichs (University of Tübingen)

- **CLARIN’s Virtual Language Observatory (VLO) under scrutiny: The VLO taskforce of the CLARIN-D centres**, a joined contribution by the German taskforce to increase the quality of the Virtual Language Observatory, presented by Susanne Haaf (BBAW), Peter Fankhauser (IDS) and Thorsten Trippel (University of Tübingen)

- **Virtual Language Observatory 3.0: What’s New?** An international cooperation, presented by Twan Goosen (CLARIN ERIC) and Thomas Eckart (University of Leipzig)

- **TeLeMaCo – A tool for the dissemination of teaching and learning materials**, a contribution by the CLARIN-D centre at the University of the Saarland by Hannah Kermes, Jörg Knappen, José Manuel Martínez, Elke Teich and Mihaela Vela

- **Your Data at the Center of Attention: A Metadata Session Profile for Multimodal Corpora**, a contribution by the users organized in the discipline-specific working group 7 (F-AG 7) by Farina Freigang, Matthias A. Priesters, Rie Nishio and Kirsten Bergmann (all Bielefeld University)

**Liaison activities with foreign CLARIN partners**

CLARIN ERIC appointed a director for user involvement who will function as a dedicated liaison officer for CLARIN as a whole. CLARIN-D also appointed a liaison coordinator to reach out to the interested scientific community outside of the CLARIN-D discipline-specific working groups (“Facharbeitsgemeinschaften”, F-AGs). These two appointments will further increase the visibility of CLARIN among potential users and related projects. The CLARIN-D and ERIC coordinators work together and harmonize their activities. CLARIN-D was involved in a European User involvement meeting on June 19th – 20th in Utrecht, The Netherlands, and at a User involvement meeting on October 24th in conjunction with the CLARIN annual meeting. CLARIN-D also participated at the following CLARIN events:


- **Joint Culture & Technology and CLARIN-D European Summer University**, Tuesday, July 22 to August 1, 2014, Leipzig

- **Exploring new ways of harvesting and generating sign language resources**, December 13 - 14, 2014, Hamburg
3 Liaison with other projects and initiatives

This section outlines recent liaison activities between CLARIN-D and other projects and initiatives such as the ESFRI project DARIAH-DE and the FP7 and Horizon2020 project EUDAT. Furthermore, as important academic centres in their various specialties, CLARIN-D members are naturally involved in a wide range of other initiatives, to which they bring their CLARIN-D experiences.

**DARIAH-DE: Joint Technical Advisory Board**

The most recent meeting of the joint Technical Advisory Board for CLARIN-D and DARIAH-DE took place at the IDS in Mannheim on February 17th, 2015. The two projects discussed areas of common interest concerning technical infrastructure, particularly persistent identifiers for resources, identity verification services, and cooperative repository services.

The board consists of the following international experts in digital humanities and computer science:

- Jonas Beskow (University of Stockholm)
- Carol Goble (University of Manchester)
- Jan Hajic (Head of the Prague CLARIN Centre)
- Ed Hovy (University of Southern California)
- Michael Lautenschlager (German Research Centre for Geosciences, Potsdam)
- Gerhard Schneider (University of Freiburg)
- Toma Tasovac (Digital Humanities Centre, Belgrade)
- Melissa Terras (University College London)
- Claire Warwick (University College London)

As part of its cooperation with DARIAH, CLARIN-D participated in the DARIAH ERIC Kick-off in Paris, November, 14th – 16th, 2014.

**EUDAT – European Data Infrastructure**

With the CLARIN Centre Tübingen as a partner in the consortium, CLARIN-D also plays a major role in the FP7 and Horizon 2020 project EUDAT (*European Data Infrastructure*), focused on building a European collaborative data infrastructure. As such, CLARIN participated at the Third EUDAT conference (September 24 - 25, 2014) and in the kick-off workshop of EUDAT2020 in Helsinki (March 24 – 26, 2015). CLARIN-D also utilizes technology provided by EUDAT and provides feedback and input to EUDAT developments.


**Research Data Alliance**

The Research Data Alliance (RDA) is a comparatively new international interdisciplinary initiative of data centres, archives and research data infrastructures. Topics of RDA are questions of technology, culture and consortia for managing and maintaining data that is created in research or that is being used for research. CLARIN-D participates with experts and use cases in RDA meetings, both nationally and internationally. In September 2014, CLARIN-D participated at the Fourth plenary Research Data Alliance in Amsterdam. In November 2014, CLARIN-D presented at the meeting of the Research Data Alliance Germany in Potsdam.

**Standardization**

CLARIN-D promotes standardization of formats and procedures for creating, describing and processing resources. This standardization is intended for long term support of research data maintenance. With the CLARIN-D centres at the IDS and in Tübingen, CLARIN-D is involved in the relevant working groups of the German Standardization Institute DIN with the subcommittee for language resources (NA 105-00-06 AA, deputy chair Andreas Witt, IDS) and Systems for Maintenance of Terminology, Knowledge and Content (NA 105-00-05 AA, deputy chair Thorsten Trippel, Tübingen). Focus in both groups is on the semantic interoperability of metadata and resource formats. CLARIN-D is also representing Germany in the international committees, especially ISO TC 37 SC 4. CLARIN-D is participating in the annual meetings and working meetings of those committees when possible, for example at the annual meeting of ISO TC 37 in Berlin June, 23rd to 27th, 2014.

CLARIN-D is also following the activities of the Text Encoding Initiative with the centres at the IDS and BBAW. This cooperation ensures the interoperability of the CLARIN-D infrastructure with those areas of the humanities implementing resources according to the TEI model.

**Liaison activities with other initiatives**

CLARIN-D members belong to the larger German, European and international academic communities. They are routinely present and active in academic conferences, workshops, and projects on many different scales. This subsection offers only a selection of liaison activities that CLARIN-D personnel have participated in as representatives of CLARIN-D over the last year.

The CLARIN-D team at the BBAW is involved in the Deutsches Textarchiv (German Text Archive); with the associated group at the University of Gießen, additional contacts to the DTA group exists. With the Gießen group, CLARIN-D reaches out to showcases for the Digital Cultural Heritage community.

CLARIN-D has also strong ties with special research centres (SFB). CLARIN-D collaborates with the Tübingen SFB 833 Emergence of meaning – the dynamics and adaptivity of linguistic structures and the Stuttgart SFB 732 Incremental Specification in Context in the development of repositories and tools. External SFBs, such as the SFB 632 and SFB 991, cooperate with
CLARIN-D in specific fields. This covers data managing with for example the SFB 632 Information structure: *The linguistic means for structuring utterances, sentences and texts* (Potsdam). In addition to data management, CLARIN assists in metadata development for example for SFB 991 *Conceptual shifts: Statistical evidence (Düsseldorf)*. This SFB also makes use of resources provided by CLARIN-D and received training in using these resources.

The Institut für Deutsche Sprache (IDS) participates in the BMBF-funded **TextGrid** project, developing sustainable, interoperable, homogenous, open source and standards-based digital research infrastructure for the humanities – a project with goals and interests that overlap with several aspects of CLARIN-D.

In the area of using resources, CLARIN-D (i.e. BBAW, IDS and UTüb) collaborates with the BMBF funded project KobRA (*Korpus-basierte linguistische Recherche und Analyse mit Hilfe von Data-Mining*, project lead: University of Mannheim) as a provider of resources.

In December 2014, CLARIN-D was supporting and hosting the International Workshop on Treebanks and Linguistic Theories in Tübingen. This workshop was attended by about 60 researchers, both from other CLARIN centres and from external projects. The infrastructure was presented with posters and information material and researchers from the CLARIN community presented applications of the resources and tools as well.

CLARIN cooperates with eHumanities projects, also funded by the BMBF. Among these projects are the GeoBib project (University of Gießen), eTRAP: electronic Text Re-use Acquisition Project (University of Göttingen), e-Identity (University of Hildesheim). Furthermore, CLARIN-D cooperates and discusses technical issues with scholars and researchers. Such discussions exist for example with a project on tools for annotating historic languages at the University of Jena. CLARIN exchanges ideas and technology with these groups and prepares for further services, i.e. the use of CLARIN web services by the projects and hosting resources for them. Other projects have been integrated into CLARIN via the F-AGs and curation projects. The creation of the F-AGs 8, 9, and 10 are indicators of this.
4 User Liaison

The CLARIN-D Liaison Coordinator, Thorsten Trippel, also regularly reaches out to Digital Humanities disciplines at the national and international level. The means of outreach are:

- Attending and presenting at conferences of the Humanities and Social Sciences
- Supporting individual projects and
- Reaching out to the relevant professional societies of individual Digital Humanities disciplines

At present, CLARIN-D focuses on those groups in the Humanities and Social Sciences showing a greater affinity to digital methods and resources. At Digital Humanities 2014 in Lausanne, Switzerland, CLARIN-D presented a special preconference workshop together with the CLARIN ERIC user involvement specialists. The topic was the use of CLARIN resources in humanities research. At the German biannual conference for computational linguistics KONVENS 2014 in Hildesheim, the CLARIN-D centres at the BAS and the University of Tübingen presented the CLARIN-D infrastructure in a dedicated session to the participants. At DHd 2015 (German speaking conference of Digital Humanities), a joint CLARIN-D and CLARIN-AT information booth was used to contact the participants. At the Leipzig summer school “Culture & Technology - The European Summer University in Digital Humanities” 2014, CLARIN had an information kiosk with changing topics to present CLARIN-D-services to the users. These activities, next to training and publications, provided valuable contacts to individuals and projects, the DHd 2014 in Graz providing about 30 new contacts expressing their interest in collaboration with CLARIN-D.
5 Further perspectives

In the coming year, activities are planned in connection with the German Society of English Philologies ("Anglistentag"), in addition to workshops and conferences as in previous years. Work package 4 will organize a dissemination workshop at the end of June, 2015 in Leipzig. This workshop will be used for reaching out to additional user groups. As in 2014, work package 8 is co-organizer of the summer school “Culture & Technology - The European Summer University in Digital Humanities”. This will provide the opportunity for additional liaison activities.

On the technical side, liaising with users will be increased using the helpdesk of work package 7 to monitor contacts with external individuals and by a redesign of the project website with portal functionalities for the CLARIN services. Also, documentation of CLARIN and CLARIN services by videos and screencasts, targeting the user groups, will be part of the liaison activities.