CLARIN-D Report
R2.2a/b:
Second Year Activity Report
Liaison with the CLARIN ERIC and Liaison with DARIAH-DE
May 2013

CLARIN, BMBF-FKZ: 01UG1120A

Deliverable: R2.2a/b
Second Year Activity Report
Liaison with the CLARIN ERI and Liaison with DARIAH-DE

Responsible: Erhard Hinrichs

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

Editors: Scott Martens, Kathrin Beck, and Thomas Zastrow
Contributors: Volker Boehlke, Alexander Geyken, Axel Herold, Hanna Kermes, Florian Schiel
Dieter Van Uytvanck, Andreas Witt, Heike Zinsmeister
1 Introduction.................................................................................................................................. 4
2 Liaisons within CLARIN............................................................................................................ 7
  2.1 Liaison activities with CLARIN ERIC.................................................................................. 7
  2.2 Liaison activities with foreign CLARIN partners .............................................................. 8
3 Liaison with DARIAH-DE ....................................................................................................... 9
  3.1 Joint Technical Advisory Board......................................................................................... 9
  3.2 Infrastructure-related working groups................................................................................ 9
  3.3 International Meetings and Dissemination.......................................................................... 9
  3.4 Individual Centres............................................................................................................... 10
4 Liaison with other initiatives.................................................................................................. 11
  4.1 BMBF-funded eHumanities projects.................................................................................. 11
  4.2 EUDAT – European Data Infrastructure ........................................................................... 12
  4.3 Liaison activities with other initiatives.............................................................................. 13
5 Conclusion and further perspectives ....................................................................................... 15
1 Introduction

CLARIN report R2.2 “Second Year Activity Report – Liaison with the CLARIN ERIC and Liaison with DARIAH-DE” documents liaison activities during the second 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) 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 Science (BBAW)
- Institute for the German Language (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 infrastructure as well as plans for the sustainable provision of digital resources and their long-term archiving.

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).

FP7 is 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

1 http://www.clarin-d.de/
2 http://cordis.europa.eu/fp7/home_en.html
3 http://ec.europa.eu/research/infrastructures/index_en.cfm?pg=esfri
1984 through 2006, and covers the period from 2007 through 2013, encompassing roughly 50 billion euros of funding activities. It is due to be replaced by the Eighth Framework Programme (FP8) 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 infrastructure 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 fulfill 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 was formally established on February 29th 2012, in a ceremony at the Dutch Ministry of Education, Culture and Science in The Hague, and is headquartered in Utrecht. It currently has nine members, eight of whom are national project organizations joining together researchers and institutions in their respective states, and one intergovernmental body:

- Austria
- Bulgaria
- Czech Republic
- Germany
- Denmark
- Estonia
- The Netherlands
- Poland
- The Nederlandse Taalunie, a joint Dutch-Flemish intergovernmental body responsible for promoting and maintaining standards for the Dutch language

---

⁴ http://ec.europa.eu/research/infrastructures/index_en.cfm?pg=eric
Norway has also joined as an observer in January 2013 with the intent to become a full member. Membership is open to European Union countries and EU-associated states as well as other participants, and 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 FP7/ESFRI roadmap project 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 European DARIAH-EU research infrastructure project.

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.

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 Liaisons within CLARIN

This section summarizes the liaison activities of CLARIN-D centres with CLARIN ERIC\(^5\) 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.

2.1 Liaison activities with CLARIN ERIC

CLARIN-D personnel 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 of the Max Planck Institute for Psycholinguistics also sits on 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.
- Heike Stadler (IDS) serves as Deputy National Content Manager.
- Thorsten Trippel of the University of Tübingen is a member of the CLARIN Office, providing support services for CLARIN ERIC.
- Andreas Witt (IDS) is the Chairman of the CLARIN ERIC Standards Committee.

Following the foundation of CLARIN ERIC, CLARIN-D signed the CLARIN Agreement (coordinated by Erhard Hinrichs and Kathrin Beck, University of Tübingen), which specifies the national contributions of CLARIN-D members to CLARIN ERIC. Germany’s CLARIN Agreement was signed in June 2012.

CLARIN-D centres and personnel participated in the CLARIN ERIC conference held in Sofia, 26-28 October 2012, both individually and in cooperation with CLARIN-NL centres:

- Erhard Hinrichs of the University of Tübingen gave the welcome address.
- Dieter Van Uytvanck presented an overview of the CLARIN technical infrastructure.
- Erhard Hinrichs and Thorsten Trippel of the University of Tübingen provided a survey of resources and tools contributed by CLARIN ERIC centres.

\(^5\) http://www.clarin.eu/
• Jens Stegmann of the University of Stuttgart gave a presentation on "Interaction, Domain Adaptation and Retraining."

• Daan Broeder from MPI Nijmegen presented reports on the CMDI metadata infrastructure and various tools to support it.

• Thomas Zastrow of the University of Tübingen gave a presentation on the Federated Content Search (FCS) specification scheme – a central component of the CLARIN-D digital infrastructure – in cooperation with Marc Kemps-Snijders of CLARIN-NL.

• Tübingen personnel Thomas Zastrow and Kathrin Beck presented several applications as potential CLARIN-D showcase projects.

Staff from several CLARIN-D centres participated in the CLARIN ERIC Federated Content Search (FCS) workshop in Mannheim on 14 June 2012, discussing issues surrounding the CLARIN-D distributed resource search system.

2.2 Liaison activities with foreign CLARIN partners

Individual CLARIN-D centres are encouraged to cooperate with foreign CLARIN partners individually and in conjunction with other CLARIN-D participants. Furthermore, there are very strong ties between CLARIN-D and CLARIN-NL centres due to past collaborations and common interests. The Max Planck Institute in Nijmegen is a centre in both CLARIN-D and CLARIN-NL, and constitutes an important connection point between the two projects.

Following groundwork established in the previous reporting period, CLARIN-D and CLARIN-NL have continued to work closely together in establishing standards for CLARIN ERIC and CLARIN as a whole, as well as moving forward to construct a joint authorization and authentication infrastructure for digital resources. CLARIN-D works closely with CLARIN-NL on ISOcat and CMDI standards, participating in the joint ISOcat coordination meeting in Utrecht on 22 August 2012, and holding a joint CMDI workshop in Nijmegen on 13 September 2012.

The Berlin-Brandenburg Academy of Sciences (BBAW) collaborates with the Institut für Corpuslinguistik und Texttechnologie (ICLTT) – a CLARIN-AT centre – on the AbAC:us – Austrian Baroque Corpus. The AbAC:us team was invited to present their results at a joint workshop in Berlin on 18 February 2013.

The BBAW also collaborates with the CLARIN-AT members as well as Swiss and South Tyrolean partners on the C4 Corpus. The C4 Corpus is an expansive corpus of regional varieties of German, consisting of texts provided by the four participating countries (Germany, Austria, Switzerland and Italy). The C4 collaborators have been jointly developing a distributed query environment for this corpus over the last years.

Dieter Van Uytvanck of MPI Nijmegen gave a presentation at the CLARINO (CLARIN Norway) metadata workshop on 5 June 2012 in Oslo. The objective of this workshop was to coordinate various initiatives in metadata cataloguing of language research data.
3 Liaison with DARIAH-DE

This section outlines recent liaison activities between CLARIN-D and DARIAH-DE, including establishing a joint advisory board on technical issues, forming cooperative infrastructure-related working groups, and holding a joint workshop.

3.1 Joint Technical Advisory Board

The joint Technical Advisory Board for CLARIN-D and DARIAH-DE had its first face-to-face meeting in Berlin on 10 April 2013. The two projects discussed a number of 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 Beskov (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)

3.2 Infrastructure-related working groups

CLARIN-D has established three joint infrastructure-related working groups in cooperation with DARIAH-DE:

- Infrastructure, standards, services, interoperability
- Curricula and training activities
- EU-Liaison

These groups are active in mediating discussions of technical standards for the two projects, in activities involving joint computing centres, and cooperative training and outreach.

3.3 International Meetings and Dissemination

CLARIN-D and DARIAH-DE organized a joint workshop at the conference Digital Humanities 2012\(^6\) in Hamburg, held 16 July 2012 and titled “Service Oriented

\(^6\) http://www.dh2012.uni-hamburg.de/
Architectures (SOAs) for the Humanities: Solutions and Impacts. Joint CLARIN-D / DARIAH workshop at Digital Humanities Conference 2012." This workshop was organized by Heike Neuroth from DARIAH-DE, and Peter Wittenburg, Erhard Hinrichs, and Thomas Zastrow from CLARIN-D.

3.4 Individual Centres

The Berlin-Brandenburg Academy of Sciences and Humanities (BBAW) is a member of both CLARIN-D and DARIAH-DE, and therefore serves as an important point of contact between the two projects. The BBAW’s IT-coordination steering committee Telota meets regularly and mediates a great deal of information exchange between the two projects, particularly through the liaison activities of Alexander Geyken, who is CLARIN-D’s formal representative on the committee.

Saarland University also cooperates closely with DARIAH-DE centres in the DARIAH-DE work package AP2: Research and Learning. It has taken part in DARIAH-DE workshops for this workpackage on May 25, 2012 in Würzburg and December 10, 2012 in Darmstadt.

---

7 http://www.dh2012.uni-hamburg.de/conference/programme/pre-conference-workshops/
4  Liaison with other initiatives

CLARIN-D and the individual CLARIN centres participate in a variety of other research initiatives. CLARIN-D centres are involved in many national eHumanities projects, generally funded by the German Federal Ministry of Research and Education (BMBF). CLARIN-D also plays a major role in the ESFRI/FP7 project EUDAT (*European Data Infrastructure*), focused on building a European collaborative data infrastructure. Furthermore, as important academic centres in their various specialties, CLARIN-D members are naturally involved in a variety of other initiatives, to which they bring their CLARIN-D experiences.

4.1  BMBF-funded eHumanities projects

In 2011, the German Federal Ministry of Research and Education (BMBF) launched a call for eHumanities research and development projects in the eHumanities (*Bekanntmachung des Bundesministeriums für Bildung und Forschung von Richtlinien zur Förderung von Forschungs- und Entwicklungsvorhaben aus dem Bereich der eHumanities*). This call is motivated by the belief that new approaches to humanities research require developing new computational techniques and solutions. These projects are encouraged to bring together specialists in the humanities with partners from the computer sciences and related technical disciplines to jointly develop new research techniques. CLARIN-D members participate as partners in several projects that have been established in response to this call.

Saarland University participates in the BMBF-funded *ARGUMENTUM* project with the DFKI (German Research Centre for Artificial Intelligence) and Europäische EDV-Akademie des Rechts (EWAR). The project’s goal is the development of an intelligent information retrieval system for legal documentation that takes advantage of automatic analysis of argumentation structure.

MPI Nijmegen participates in the BMBF-funded *AUVIS* project. AUVIS deals with automatic analysis of audio and video data and event segmentation in utterances and cognition.

Prof. Dr. Henning Lobin of the University of Giessen is associated with the CLARIN-D project and the leader of the BMBF-funded *GeoBib* project. GeoBib analyses the backgrounds and content of literature published during the Holocaust in concentration camps.

The BBAW, the Institut für Deutsche Sprache (IDS) and the University of Tübingen participate in the BMBF-funded *KobRA* project, *Corpus-based linguistic research and analysis with the help of data mining*. KobRA is developing lexical data to serve as a basis for training automated processing algorithms for document clustering and for identifying new words and meanings. Saarland University also participated in the kick-off meeting for KobRA at TU Dortmund on 15-16 November 2012.

The University of Stuttgart participates in several BMBF-funded projects with other universities and institutes:

---

8  [http://www.bmbf.de/foerderungen/16466.php](http://www.bmbf.de/foerderungen/16466.php)
• **eIdentity** – *Multiple collective identities in international debates regarding war and peace since the end of the Cold War. Language technology tools and methods for the analysis of multi-lingual text in the social sciences.*

Stuttgart and CLARIN-D will assist in adapting natural language processing tools for a specialized search.

• **VisArgue** – *Why and when do arguments win? An analysis and visualization of political negotiations.*

Stuttgart and CLARIN-D will provide corpus infrastructure support and assistance in metadata construction.

• **ePoetics** – *Corpus development and visualization for German poetry study (1770-1960) for “algorithmic criticism.”*

Stuttgart and CLARIN-D will provide support in adapting natural language processing tools to historical and non-standard language, as well as corpus management and metadata support.

The CLARIN-D centre at the University of Leipzig is also a participant in two BMBF-funded projects with other German research centres:

• **ePol** – *The use of neoliberal arguments in German federal politics, 1949-2011.*

• **eTraces** – *Winged words, quotations and our cultural heritage. Studying temporal traces and interconnecting relations of text passages in German language novels from between 1500 and 1900, as well as social science texts created since 1909.*

At the eHumanities workshop in Leipzig, two BMBF-funded projects, ARGUMENTUM and SeNeReKo requested advice from CLARIN-D concerning annotation tools for digitized texts. Both projects were brought into contact with Prof. Dr. Chris Biemann of the University of Darmstadt, who is responsible for developing the WebAnno annotation tool as part of CLARIN-D.

### 4.2 EUDAT – European Data Infrastructure

The EUDAT⁹ project is building a collaborative data infrastructure for researchers to share data with colleagues from all disciplines, helping them to collaborate effectively. The CLARIN community acts as the linguistics, humanities and social science user group participating in EUDAT.

MPI Nijmegen and the University of Tübingen are partners in both CLARIN-D and EUDAT, and both centres contribute to various EUDAT work packages. They are particularly involved in the following areas:

• **Sustainability.**

  The continuing availability and relevance of shared services and resources beyond the immediate project timeframe is a goal of both CLARIN-D and EUDAT, and experience developed in one project is immediately applicable to the other.

• **Workflow development.**

  Tübingen contributes web services and know-how from the WebLicht

---

⁹ [http://www.eudat.eu/]
environment, developed in the context of D-SPIN and CLARIN-D, towards building a generic workflow framework for EUDAT.

- Safe replication for data repositories. Nijmegen and Tübingen both contribute knowledge from their CLARIN-D experience in building repositories towards EUDAT's goals.

### 4.3 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 BBAW participated in a special workshop at the conference *Historische Textkorpora für die Geistes- und Sozialwissenschaften. Fragestellungen und Nutzungsperspektiven* (18–19 February 2013) in cooperation with the Deutsches Textarchiv (German Text Archive). By means of an integrated workshop a brief introduction to CLARIN-D was presented.

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.

The University of Tübingen is involved in the SFB (Sonderforschungsbereich) project SFB 833 *Emergence of meaning – the dynamics and adaptivity of linguistic structures*, and provides that project with natural language processing infrastructure developed as part of CLARIN-D.

The University of Stuttgart is involved in SFB 732 *Incremental Specification in Context*, providing support in metadata development; and SFB 991 *Conceptual shifts: Statistical evidence*, offering advice on annotation, corpus infrastructure and metadata.

The University of Stuttgart also contributes advice on corpus infrastructure and metadata to the German Research Centre project *SynAlm: Syntax of the Alemannic dialects*, and participates in the German Research Centre-funded Netzwerk Kobalt-DaF using CLARIN-D infrastructure to support corpus research in German second language acquisition studies.

Saarland University participated in the workshop *New developments in software tools for corpus construction and analysis* at the University of Bologna. The workshop focused on Italian initiatives to collect software tools for corpus construction and analysis with an emphasis on parallel corpora. Saarland University was invited to this event as the CLARIN-D centre with particular expertise in parallel corpora.

MPI Nijmegen has been in contact with a number of international projects with interests related to those of CLARIN. Dieter van Uytvanck attended a meeting with The European Library – a digital resource federation of European library federations and related projects. The MPI also participated in the METAshare metadata workshop at LREC in Istanbul in May 2012 and in the FIM4R (Federated Identity Management for
Researchers) workshop in Villigen, Switzerland on 20 March 2013.

The Institut für Deutsche Sprache participates in European and international standards bodies, including ISO/TC 37/SC 4, which provides standards for language resource management, and the Text Encoding Initiative (TEI) special interest group for linguistic data standards (TEI for Linguists SIG). It is also involved in the European Commission’s Text and Data Mining Working Group (WG4), and the Research Data Alliance, particularly its Legal Issues working group.

The University of Leipzig is in close contact with the Perseus Project at Tufts University (NY, USA), specialized in digitizing and visualizing classical, medieval and early American texts for humanities research.

The Bavarian Archive for Speech Signals (BAS) has been very active in maintaining contact with Australian researchers, and has established a work group for “Australian eHumanities Infrastructures and CLARIN” under the leadership of Professor Steve Cassidy of Macquarie University. BAS involvement includes sending Dr. Florian Schiel to the University of Melbourne in early 2013, and reciprocally hosting Dr. Felicity Cox of Macquarie University at Munich in May 2013. It has also been in contact with a variety of external research groups through the promotion and use of WebMAUS, a CLARIN-D and BAS-developed web application for annotating recorded speech.
5 Conclusion and further perspectives

During the past year, CLARIN-D has extended the basis for interaction established in the previous year with German national projects – particularly eHumanities projects funded by the BMBF as well as the ESFRI/FP7 project DARIAH-DE – and internationally with other national CLARIN members through direct contact and through CLARIN ERIC. It has also worked to raise the profile of CLARIN-D in international contexts and cooperates with a variety of other initiatives.

CLARIN-D is making rapid progress in implementing digital language and textual infrastructure that is already being used by German and international research projects through the extensive liaison efforts of CLARIN-D and its member centres.

CLARIN ERIC will in the immediate future appoint a director for user involvement who will function as a dedicated liaison officer for CLARIN as a whole. CLARIN-D also plans to employ a liaison coordinator starting in 2014. These two appointments will further increase the visibility of CLARIN among potential users and related projects.