

Antonina Werthmann/Andreas Witt/Jutta Bopp (Mannheim)  
**Verbundprojekt CLARIAH-DE –  
Eine nachhaltige Forschungsinfrastruktur  
für die Geistes-, Kultur- und  
Sozialwissenschaften**

**Abstract:** Das vom BMBF geförderte Verbundprojekt CLARIAH-DE, an dem über 25 Partnerinstitutionen mitwirken, unter ihnen auch das IDS, hat zum Ziel, mit der Entwicklung einer Forschungsinfrastruktur zahlreiche Angebote zur Verfügung zu stellen, die die Bedingungen der Forschungsarbeit mit digitalen Werkzeugen, Diensten sowie umfangreichen Datenbeständen im Bereich der geisteswissenschaftlichen Forschung und benachbarter Disziplinen verbessern. Die in CLARIAH-DE entwickelte Infrastruktur bietet den Forschenden Unterstützung bei der Analyse und Aufbereitung von Sprachdaten für linguistische Untersuchungen in unterschiedlichsten Anwendungskontexten und leistet somit einen Beitrag zur Entwicklung der NFDI.

2019 startete das CLARIAH-Projekt<sup>1</sup> als Zusammenschluss der zwei Forschungsinfrastrukturverbünde CLARIN-D<sup>2</sup> und DARIAH-DE<sup>3</sup>. CLARIN-D ist der deutsche Beitrag zur *Common Language Resources and Technology Infrastructure* (CLARIN), die als *European Research Infrastructure Consortium* (ERIC) ein Netzwerk aus 20 Ländern neben weiteren Partnern bildet. DARIAH-DE ist der deutsche Partner der *Digital Research Infrastructure for the Arts and Humanities* (DARIAH-EU), die mit 19 Ländern und weiteren Partnern ebenfalls ein ERIC auf europäischer Ebene ist. Sowohl CLARIN-D als auch DARIAH-DE bestehen bereits seit über zehn Jahren und sind in der geisteswissenschaftlichen Forschung sowie in weiteren Disziplinen etabliert. Gemeinsam stellen sie digitale Forschungsdaten und Werkzeuge zu deren Bearbeitung zur Verfügung, bieten regelmäßige Schulungsangebote sowie webbasierte Hilfen im Umgang mit diesen und unterstützen dadurch die mit digitalen Ressourcen und Methoden arbeitenden Wissenschaftlerinnen und Wissenschaftler in Forschung und Lehre.

---

1 [www.clariah.de](http://www.clariah.de) (Stand: 19.10.2020)

2 [www.clarin-d.net/de/](http://www.clarin-d.net/de/) (Stand: 19.10.2020)

3 [de.dariah.eu/web/guest/startseite](http://de.dariah.eu/web/guest/startseite) (Stand: 19.10.2020)

CLARIAH-DE wird vom Bundesministerium für Bildung und Forschung (BMBF) für zwei Jahre gefördert. Das Hauptziel für die Zusammenführung der beiden Forschungsinfrastrukturverbünde besteht darin, den Nutzerinnen und Nutzern eine Möglichkeit anzubieten, auf komfortable Weise über einen gemeinsamen Zugang auf alle Dienste und Ressourcen zuzugreifen, unabhängig davon, von welchem der zwei Verbünde sie angeboten werden. Das gesamte Portfolio von CLARIAH-DE kann über den bestehenden eduGAIN-Zugang der jeweiligen Heimatinstitutionen oder mit einem CLARIN- oder DARIAH-Account einfach verwendet werden. Die digitalen Werkzeuge, Dienste und Datenbestände werden gemäß der FAIR-Prinzipien (FAIR = findable, accessible, interoperable, re-usable) auffindbar, zugänglich, interoperabel und nachnutzbar gemacht. Durch die Verschmelzung von CLARIN-D und DARIAH-DE werden die ursprünglich voneinander getrennt entwickelten Angebote, die aus gemeinsamen technischen Komponenten sowie abgestimmten Verfahren bestehen, in eine einzige Forschungsinfrastruktur integriert und leisten somit einen Beitrag zur Entwicklung der Nationalen Forschungsdateninfrastruktur (NFDI).

Bei der Zusammenführung digitaler Datenbestände steht CLARIAH-DE vor der Aufgabe, die von CLARIN-D und DARIAH-DE geschaffenen Angebote aufeinander abzustimmen, in eine neue Struktur zu integrieren, weiterzuentwickeln und gemeinsam weiter zu betreiben. Eine der Herausforderungen stellt dabei die Heterogenität der Daten und Werkzeuge dar. Aufbauend auf früheren Kooperationen und Abstimmungsprozessen werden beide Infrastrukturen die hierzu notwendigen Prozesse etablieren, Standards und Verfahren für die Erstellung, Aufbereitung sowie Archive gestalten und gleichzeitig die fachliche Breite des Angebots erhalten.



**Abb. 1:** Zusammenführung der Forschungsinfrastrukturverbünde CLARIN-D und DARIAH-DE

Durch die Zusammenführung der Forschungsinfrastrukturverbünde CLARIN-D und DARIAH-DE bietet CLARIAH-DE den Forscherinnen und Forschern eine komplexe digitale Infrastruktur mit einer Vernetzung von Anwendungen und Koordination der technischen Entwicklungen (siehe Abb. 1). Die vielseitigen Angebote reichen von virtuellen Forschungsumgebungen bis hin zu Werkzeugen, mit denen geeignete Daten gefunden, annotiert und editiert werden können, unabhängig davon, wo oder von wem die jeweiligen Angebote ursprünglich aufgebaut bzw. bereitgestellt wurden. Darüber hinaus finden sich zahlreiche Anwendungen zur Datenvisualisierung und -analyse. Die CLARIAH-DE-Angebote beschränken sich allerdings nicht auf die Bereitstellung von Diensten und Werkzeugen, sondern stellen auch ein Sortiment an Informationen zu Lernmaterialien im Umgang mit diesen zur Verfügung. Innerhalb der Projektlaufzeit von zwei Jahren werden zahlreiche Workshops und Kurse zu Methoden und Technologien zur Kompetenzvermittlung angeboten sowie Summer Schools (mit)organisiert. Darüber hinaus werden Stipendien zur Nachwuchsförderung vergeben. Durch die Bereitstellung der Angebote zielt CLARIAH-DE darauf ab, die Community zu erweitern und zu stärken. Perspektivisch soll die Nutzbarkeit der Angebote nicht nur von Digital Humanities, sondern auch von anderen Forschungsgemeinschaften erschlossen und somit die Nutzerbasis erweitert werden. Bei Fragen und Problemen rund um CLARIAH-DE steht den Nutzerinnen und Nutzern außerdem ein zentraler Helpdesk (E-Mail: [support@clariah.de](mailto:support@clariah.de)) zur Verfügung.

25 universitäre und außeruniversitäre Institutionen bestehend aus zehn Partnern, 13 assoziierten Partnern und zwei Unterauftragnehmern arbeiten an diesem Verbundprojekt zusammen (siehe Abb. 2), daneben Fachleute von vielen weiteren akademischen Institutionen. Gemeinsam bringen sie eine komplementäre Expertise und zahlreiche Ressourcen und Werkzeuge zur Förderung einer kooperativen, digital gestützten Forschung mit. Das Gesamtvorhaben ist in sechs Arbeitspakete (AP 1–6) aufgeteilt. Die Gesamtkoordination liegt gemeinschaftlich bei der Niedersächsischen Staats- und Universitätsbibliothek Göttingen und der Eberhard Karls Universität Tübingen.

Über CLARIN-D beteiligt sich das IDS am Projekt mit der Weiterentwicklung von Schnittstellen zwischen den Plattformen für schriftliche und mündliche Korpora (KorAP bzw. DGD) und der föderierten Inhaltssuche (Federated Content Search, FCS), der Einbindung der KorAP-Resultatsanalyse in die CLARIN-Infrastruktur sowie der Implementierung der CLARIN-konformen Virtual Collection Registry, die als Basis für die Definition und die persistente Referenzierung virtueller Korpora dienen soll. Außerdem betreut das IDS den Legal Helpdesk, der Unterstützung bei rechtlichen und ethischen Fragen zur Aufbereitung und Verbreitung von Sprachressourcen bietet.



Abb. 2: Am Verbundprojekt CLARIAH-DE beteiligte Institutionen<sup>4</sup>

In CLARIAH-DE ist das Leibniz-Institut für Deutsche Sprache (IDS) als Co-Taskleader für AP 3 *Kompetenzvermittlung und Nachwuchsförderung* mitverantwortlich. Darüber hinaus beteiligt sich das IDS an den Arbeiten in den Arbeitspaketen *Forschungsdaten, Standards und Verfahren* (AP 1), *Werkzeuge und virtuelle Forschungsumgebungen* (AP 2), *Technische Vernetzung und Koordination der technischen Entwicklungen* (AP 4) und *Community-Engagement: Outreach/Dissemination und Liaison* (AP 5).

<sup>4</sup> Siehe: [www.clariah.de/ueber-uns/partner](http://www.clariah.de/ueber-uns/partner) (Stand: 19.10.2020).