
Welche Unterstützungsangebote benötigen Disziplinen für die systematische Verankerung von E-Science in der universitären Forschung und Lehre? Perspektiven und Anforderungen am Beispiel der Germanistischen Linguistik

Michael Beißwenger¹, Hubert Klüpfel², Ania López³ und Stephanie Rehwald³

¹Institut für Germanistik, Universität Duisburg-Essen;

²Referent des CIO Universität Duisburg-Essen;

³Universitätsbibliothek, Universität Duisburg-Essen

In diesem Beitrag diskutieren wir Anforderungen und Perspektiven für die Unterstützung von E-Science an Universitäten. Am Beispiel der Germanistischen Linguistik skizzieren wir die Bedeutung der Arbeit mit digitalen Daten und -infrastrukturen in Forschung und Lehre und leiten daraus Anforderungen ab, die sich bei der Integration entsprechender Ressourcen in den Forschungsprozess, in die Lehre und in die Ausbildung des wissenschaftlichen Nachwuchses ergeben. Wir geben einen Überblick über außeruniversitäre Initiativen und Netzwerke, die für die Disziplin relevante Infrastrukturen und Services bereitstellen oder solche entwickeln. Diese Angebote können die in dieser und anderen Disziplinen bestehenden Anforderungen bislang aber nur zu einem Teil abdecken. Für die Universitäten stellt sich die Frage, inwieweit sie – ergänzend zu vorhandenen Initiativen und Netzwerken – die Dissemination und Weiterentwicklung von E-Science als Institution aktiv mitgestalten möchten und welche Unterstützungsmaßnahmen sie dazu im Rahmen ihrer Aufgaben bereitstellen können. Am Beispiel der Universität Duisburg-Essen zeigen wir, wie Unterstützungsmaßnahmen, insbesondere auch mit Blick auf die gegenwärtig in Aufbau befindlichen Forschungsdateninfrastrukturen auf nationaler (NFDI) und auf Länderebene, konzipiert und weiter ausgebaut werden können.

1. Einleitung

In diesem Beitrag diskutieren wir Perspektiven und Strukturen für die Unterstützung von E-Science an der Universität Duisburg-Essen (UDE). Am Beispiel der Germanistischen Linguistik und mit Bezug zu konkreten Projekte beschreiben wir die Anforderungen, die sich bei der Einbindung digitaler Forschungsressourcen und -methoden in Forschungsprozesse und in die Lehre stellen. Wir geben einen groben Überblick, welche außeruniversitären Initiativen und Unterstützungsangebote existieren, auf die Forschende und Lehrende für die Arbeit mit digitalen Ressourcen zurückgreifen können. Wir skizzieren, welche

Unterstützungsangebote an der Universität Duisburg-Essen – wie vermutlich auch an anderen Hochschulen – bereits vorhanden sind und wie diese die Anforderungen aus dem Fach bedienen können. Davon ausgehend diskutieren wir, welche Rolle die Universität, ergänzend und komplementär zu existierenden außeruniversitären Angeboten, in Bezug auf die Weiterentwicklung von E-Science in der universitären Forschung und Lehre einnehmen könnte. Speziell im Zusammenhang mit der im Aufbau befindlichen Nationalen Forschungsdateninfrastruktur (NFDI) und dazu korrespondierenden Initiativen auf Landesebene werden die Universitäten nicht umhinkommen, ihr Selbstverständnis und ihre Strategie in Bezug auf die Förderung und Verbreiterung von E-Science in den Disziplinen zu bestimmen.

2. Bedeutung von digitalen Forschungsressourcen für die Linguistik

Die Germanistische Linguistik greifen wir in unserem Beitrag als ein Paradebeispiel für eine Disziplin heraus, in der digitale Forschungsressourcen und Infrastrukturen die Möglichkeiten und Methoden empirischer Forschung in den letzten zwei Jahrzehnten substantiell verändert haben. Die Arbeit mit digitalen Ressourcen und Werkzeugen ermöglicht es, wissenschaftliche Hypothesen mit computergestützten Methoden an dokumentierten Stichproben zum Sprachgebrauch zu erkunden sowie qualitativ und quantitativ zu überprüfen. Digitale Infrastrukturen, über die diese Ressourcen und Werkzeuge genutzt werden können und die unter Beteiligung von Forscher*innen im Fach weiter ausgebaut werden, bilden daher „eine wichtige und in der Zukunft noch wichtigere Säule für das Fach“ (Hinrichs 2018: 47).

Als Datengrundlage für linguistische Forschungen spielen u.a. sog. *Korpora* eine wichtige Rolle. Darunter versteht man Sammlungen authentischer Sprachdaten (mündlich, schriftlich, multimodal), die für Zwecke der linguistischen Analyse aufbereitet sind. Typische Aufbereitungsschritte sind z. B. die Anreicherung der Daten um *Metadaten*, die für die Analyse benötigt werden (bei Texten z. B. Angaben zum Autor, zur Textsorte, zum Erscheinungsort und -jahr, zu möglichen Vor- und Nebenversionen; bei Gesprächen und Daten aus internetbasierter Kommunikation z. B. Angaben zu den Interaktionsbeteiligten und zu ihren sozialen Beziehungen, zu ihrem Alter und Bildungsgrad, zu verwendeten Sprachen, zum Zweck und zur Vorgeschichte der Interaktion usw.), sowie das Hinzufügen von linguistischen Annotationen (z. B. Wortart, Flexionsform, syntaktische Struktur) zu den enthaltenen Wörtern, syntaktischen Aufbaueinheiten und Sätzen. Mit Annotationen wird es möglich, bei der Korpus-Recherche und -analyse über die reine Volltextsuche hinaus nach sprachlichen Mustern und Strukturen zu suchen (Lemnitzer & Zinsmeister 2015; Lüdeling/Kytö 2008/2009).

Eine besondere Rolle spielen sog. *Referenzkorpora*. Das sind Korpora, die nicht nur in der Eigenforschung des Korpus-Erstellers eine Rolle spielen, sondern die der Scientific Community als Grundlage für beliebige sprachbezogene Forschungen bereitgestellt werden. Forschung, die auf Grundlage solcher Korpora durchgeführt wird, ist nachprüfbar, weil die Daten offen zugänglich sind. Eine Vernetzung der Forschungsergebnisse mit

den verwendeten Ressourcen wird möglich. Dadurch, dass unterschiedliche Forschungsfragen auf denselben Referenzkorpora durchgeführt werden, wird es möglich, Ergebnisse zu unterschiedlichen Forschungsfragen über den Bezug auf dieselben Daten zueinander in Beziehung zu setzen und sie zu vergleichen. Dadurch können neue Forschungsfragen entstehen.

Große, aufbereitete Referenzkorpora und Korpus-Sammlungen, wie sie z. B. vom Leibniz-Institut für Deutsche Sprache (IDS), Mannheim, mit dem Deutschen Referenzkorpus DeReKo (Lüngen 2017) und der Datenbank Gesprochenes Deutsch (Schmidt 2017) oder von der Berlin-Brandenburgischen Akademie der Wissenschaften (BBAW) über die Korpusplattform des Projekts www.dwds.de (Geyken et al. 2017) angeboten werden und die über Online-Schnittstellen abgefragt werden können, bieten Wissenschaftler*innen für viele Domänen des Sprachgebrauchs bereits eine gute Ressourcenlage. Für andere, noch unzureichend abgedeckte Domänen wird die Ressourcenlandschaft in diversen Projekten weiter ausgebaut. Eine solche Domäne ist aktuell die Sprachverwendung in Social-Media-Anwendungen bzw. in der internetbasierten Kommunikation (Beißwenger et al. 2017, Beißwenger 2018). Neben Korpora spielen im Bereich der Linguistik aber auch noch andere digitale Sprachressourcen (DSR) eine Rolle – zum Beispiel digitale Wörterbücher und Nachschlagewerke zur Sprache (z.B. Grammatiken), Sammlungen historischer Texte, lexikalisch-semantische Netze oder Werkzeuge für die automatische Sprachanalyse (Tokenisierer, Part-of-Speech-Tagger, Lemmatisierer, syntaktische Parser u.a.). Als sehr produktiv für den Aufbau und die Aufbereitung von Sprachkorpora und anderen digitalen Sprachressourcen hat sich die Zusammenarbeit mit dem Bereich Sprachtechnologie/Computerlinguistik erwiesen, der Verfahren für die automatische linguistische Analyse und für die Annotation von Sprachdaten entwickelt. Nationale und internationale Initiativen unterstützen den Ausbau und die Bereitstellung von Sprachressourceninfrastrukturen im Bereich der Linguistik und der Geisteswissenschaften.

3. Welche Anforderungen stellen sich in der Disziplin?

In diesem Abschnitt skizzieren wir Anforderungen an die Arbeit mit digitalen Sprachressourcen (DSR) in der universitären Forschung und Lehre im Fach Germanistik.¹ Dazu unterscheiden wir zunächst sechs verschiedene Perspektiven, unter denen DSR für Wissenschaftler*innen im Fach im Zusammenhang mit ihren Aufgaben in Forschung und Lehre relevant werden. Welche dieser Perspektiven für den oder die einzelne*n Wissenschaftler*in tatsächlich praxisrelevant sind, hängt ab vom Stellenprofil und von individuellen thematischen Schwerpunkten. Begreift man die Arbeit mit DSR und mit darauf bezogenen Methoden als einen wichtigen Baustein einer modernen, empirisch ausgerichteten Linguistik, so darf angenommen werden, dass für einzelne Wissenschaftler*innen jeweils mehrere der Perspektiven in der Forschungs- und Lehrpraxis eine Rolle spielen. Für an der Universität tätige Linguist*innen ist der Umgang mit digitalen Sprachressourcen (DSR) unter unterschiedlichen Perspektiven relevant:

1. Als Forschende nutzen sie DSR als Werkzeuge im eigenen Forschungsprozess.

2. Als Lehrende vermitteln sie Möglichkeiten der Arbeit mit DSR an Studierende, um diese mit modernen empirischen Methoden des linguistischen Arbeitens vertraut zu machen.
3. Als Hochschullehrer*innen bilden sie wissenschaftlichen Nachwuchs aus, für dessen Karriere ein reflektierter Umgang mit DSR unerlässlich ist.
4. Als empirisch Forschende leiten sie Forschungsprojekte, in deren Rahmen Sprachdaten erhoben werden, um diese linguistisch zu analysieren. Forschungsförderer wie z. B. die DFG erwarten, dass in Projekten erhobene Forschungsdaten bzw. Sprachkorpora nachnutzbar vorgehalten werden, und geben dafür datentechnische und rechtliche Standards vor (vgl. DFG 2015a, 2015b).
5. Ggf. sind sie auch an Forschungs- und Entwicklungsprojekten beteiligt, deren genuines Ziel im Aufbau von DSR besteht, die eine Lücke in der Korpuslandschaft schließen und der Scientific Community als Forschungsressourcen bereitgestellt werden sollen.
6. Als Akteur*innen im Bereich ‚Digital Humanities‘ und/oder als Mitglied von Fachgesellschaften beteiligen sie sich ggf. an der Weiterentwicklung von Standards für die Arbeit mit DSR und am Ausbau entsprechender Infrastrukturen.

Diese Perspektiven dürften sich *mutatis mutandis* auch auf die Arbeit mit digitalen Ressourcen in anderen geisteswissenschaftlichen Disziplinen übertragen lassen. Je nach Perspektive ergeben sich dabei unterschiedliche Anforderungen an Unterstützungsangebote, die benötigt werden, wenn man die Weiterentwicklung der damit beschriebenen Aktivitäten in Forschung und Lehre als ein strategisches Entwicklungsziel des Faches und der das Fach einbettenden Institutionen (Universität, Fachgesellschaften, Forschungsdateninfrastrukturen) begreift. Die folgenden Anforderungen lassen sich – als erweiterbare Liste – formulieren; der Fokus liegt auf der Arbeit mit Korpora.

1. Nutzung von DSR im Forschungsprozess: Benötigt werden Schnittstellen und Werkzeuge, anhand derer existierende DSR flexibel für die Nutzung in unterschiedlichen linguistischen Forschungskontexten angepasst werden können. Aus DSR für Forschungszwecke extrahierte Daten sollten, auch nach weiterer Aufbereitung, jederzeit auf die entsprechenden Teile der Ausgangsressource rückbeziehbar sein. Projektspezifisch vorgenommene Erweiterungen von Ausschnitten bestehender Ressourcen (z. B. zusätzliche linguistische Annotationen) sollten so zur Verfügung gestellt werden können, dass sie für die Scientific Community als Erweiterungen der Ausgangsressource nutzbar sind.
2. Arbeit mit DSR in der universitären Lehre: DSR sollten mit Nutzerschnittstellen ausgestattet sein, die es auch Semi-Experten der Domäne nach entsprechender Einarbeitung erlauben, in vollem Umfang von den Abfrage- und Auswertungsmöglichkeiten Gebrauch zu machen. Um die Potenziale von DSR für die empirische Forschung praxisnah zu vermitteln, ist die Bereitstellung didaktisch aufbereiteter Praxisbeispiele für Forschungsprojekte auf unterschiedlichen Niveaustufen wünschenswert.

3. Ausbildung des wissenschaftlichen Nachwuchses: Wünschenswert sind regelmäßige Weiterbildungsangebote, z. B. auf Ebene von Fächergruppen oder Fakultäten, zu existierenden DSR und zu Methoden der Ressourcen-gestützten linguistischen Forschung, in denen gängige Werkzeuge, Standards und Verfahren für die Modellierung und Annotation von Textdaten sowie für die Abfrage und computergestützte Auswertung von Korpora vorgestellt werden. Solche Weiterbildungen sind nur als fachnah konzipierte Angebote sinnvoll denkbar. Expertise zu Standards, Werkzeugen und Infrastrukturen im Bereich Digital Humanities sowie personelle Ressourcen für deren Vermittlung ist für eine nachhaltige Verankerung solcher Angebote unabdingbar.
4. Selbst (z. B. im Rahmen von geförderten Projekten) erhobene Forschungsdaten nachnutzbar aufbereiten und im Rahmen etablierter Forschungsinfrastrukturen bereitstellen: Hier besteht Bedarf an Beratung und Unterstützung, um in Projekten erhobene Daten in Repräsentationsformate zu konvertieren, die für die Langzeitarchivierung in Repositorien und ggf. für die Bereitstellung als DSR in existierenden Forschungsinfrastrukturen als Standards etabliert sind. Erfahrungsgemäß ist hierzu Beratung bereits in der Konzeptionsphase der Projekte erforderlich, damit idealerweise schon vor der Erhebung der Projektdaten Modellierungsentscheidungen getroffen werden können, die eine nachhaltige Repräsentation erlauben. Auch rechtliche und forschungsethische Fragen sind entsprechend frühzeitig im Erhebungsprozess zu behandeln, wenn nach Abschluss des Projekts eine Archivierung und/oder Weitergabe an Dritte vorgesehen (oder vom Projektträger erwünscht) ist. Um die Integration von Projektdaten in existierende übergreifende Infrastrukturen (z. B. CLARIN) zu organisieren, sollten Projektleiter mit den entsprechenden Zentren in Kontakt gebracht und sollten technische Fragen der Datenübergabe sowie der vor- oder nachgängigen Datenaufbereitung durch eine Supportstelle mit Expertise im Bereich Digital Humanities unterstützt werden.
5. Aufbau und Erweiterung des DSR-Ressourcenlandschaft: Der Aufbau neuer DSR und die Beteiligung am Ausbau von Forschungsdateninfrastrukturen sowie die damit zusammenhängenden konzeptionellen und technischen Aufgaben sollten als Forschungsleistungen – beispielsweise im Rahmen von Ziel- und Leistungsvereinbarungen – gewürdigt werden. Technische Unterstützungsmaßnahmen – z. B. die Bereitstellung projektspezifisch eingerichteter Server – und konzeptionelle Services – z. B. Unterstützung bei der Entwicklung geeigneter Datenhaltungsstrukturen – ist in der Entwicklungsphase neuer DSR erforderlich, da in den Geisteswissenschaften typischerweise kein Personal für solche Aufgaben vorhanden ist. Nach Abschluss der Entwicklungsarbeiten, der Datenerhebung und Aufbereitung sollten die resultierenden DSR dann in übergreifende Forschungsdateninfrastrukturen überführt werden (siehe 4.).
6. Beteiligung an der Weiterentwicklung von Standards: Die Weiterentwicklung von Standards für die Arbeit mit DSR und am Ausbau entsprechender Infrastrukturen sollte als Forschungsleistung – beispielsweise im Rahmen von Ziel- und Leistungsvereinbarungen – gewürdigt werden.

4. Existierende Infrastrukturen und Netzwerke außerhalb der Universitäten

Die Dissemination der Nutzung von DSR und von Werkzeugen für die computergestützte Analyse von DSR für die empirische linguistische Forschung und für eine zeitgemäße wissenschaftliche Ausbildung von Studierenden und von wissenschaftlichem Nachwuchs wird u.a. in Initiativen wie CLARIN-EU¹ bzw. CLARIN-D² und DARIAH³ vorangetrieben. Auch verschiedene Fachgesellschaften widmen sich dem Thema mit eigenen Sektionen – so beispielsweise die Sektion „Computerlinguistik“ in der deutschen Gesellschaft für Sprachwissenschaft (DGfS)⁴, der Forschungsfokus „Digitale Infrastrukturen für die Angewandte Linguistik“ (GAL-DIAL) in der Gesellschaft für Angewandte Linguistik (GAL e.V.)⁵ sowie verschiedene Arbeitskreise in der Gesellschaft für Sprachtechnologie und Computerlinguistik (GSCL e.V.)⁶.

Der Förderung einer Vernetzung zwischen Sprachressourcenanbietern und linguistischen Nutzer*innen verpflichtet sind diverse Aktivitäten (internationale Workshops, Master Classes, Summer Schools) von CLARIN-ERIC⁷ zum Thema „User Involvement“. Mit der Entwicklung eines Curriculums für die „Digitalen Geisteswissenschaften“ beschäftigt sich u. a. die AG „Referenzcurriculum Digital Humanities“ in der Fachgesellschaft Digital Humanities im deutschsprachigen Raum (DHd)⁸.

Das deutsche CLARIN-Verbundprojekt CLARIN-D ist organisiert in Zentren als Pfeilern einer aufzubauenen, im europäischen Kontext eingebundenen, nationalen Sprachressourcen- und Forschungsinfrastruktur. Die Zentren stellen Forschenden DSR und darauf bezogene Services bereit und kuratieren DSR in Hinblick auf die Langzeitarchivierung und Bereitstellung für die Scientific Community. Mit der Förderung der „Nationalen Forschungsdateninfrastruktur“ (NFDI) durch Bund und Länder wird eine Bündelung und Koordination bestehender und neuer Aktivitäten im Forschungsdatenmanagement angestrebt. Wie in Abbildung 1 gezeigt, liegt die NFDI quer zu den vorhandenen Strukturen und folgt dem Leitgedanken, Konsortien zu bilden, die verschiedene Säulen des Wissenschaftssystems miteinander vernetzen. Wie bei dem Aufbau der European Open Science

¹ CLARIN: Common Language Resources and Technology Infrastructure. <https://clarin.eu>, besucht: 26.04.2019.

² CLARIN-D: Deutsches Konsortium im Rahmen von CLARIN-EU: <https://www.clarin-d.net>, besucht: 05.04.2019.

³ DARIAH: Digital Research Infrastructure for the Arts and Humanities. <https://dariah.eu>, besucht: 26.04.2019.

⁴ Deutsche Gesellschaft für Sprachwissenschaft / Sektion Computerlinguistik, 2019. <https://dgfs.de/de/c1/>, besucht: 26.04.2019.

⁵ Gesellschaft für Angewandte Linguistik / Forschungsfokus Digitale Infrastrukturen für die Angewandte Linguistik (GAL-DIAL): <https://gal-ev.de/gal-forschungsfokus-digitale-infrastrukturen-fuer-die-angewandte-linguistik-gal-dial/>, besucht: 26.04.2019.

⁶ Arbeitskreise in der Gesellschaft für Sprachtechnologie und Computerlinguistik: <https://gscl.org/sigs>, besucht: 04.05.2019.

⁷ CLARIN-ERIC: <https://www.clarin.eu/>, besucht: 04.05.2019.

⁸ DHd-AG Referenzcurriculum Digital Humanities: <https://dig-hum.de/ag-referenzcurriculum-digital-humanities>, besucht: 26.04.2019.

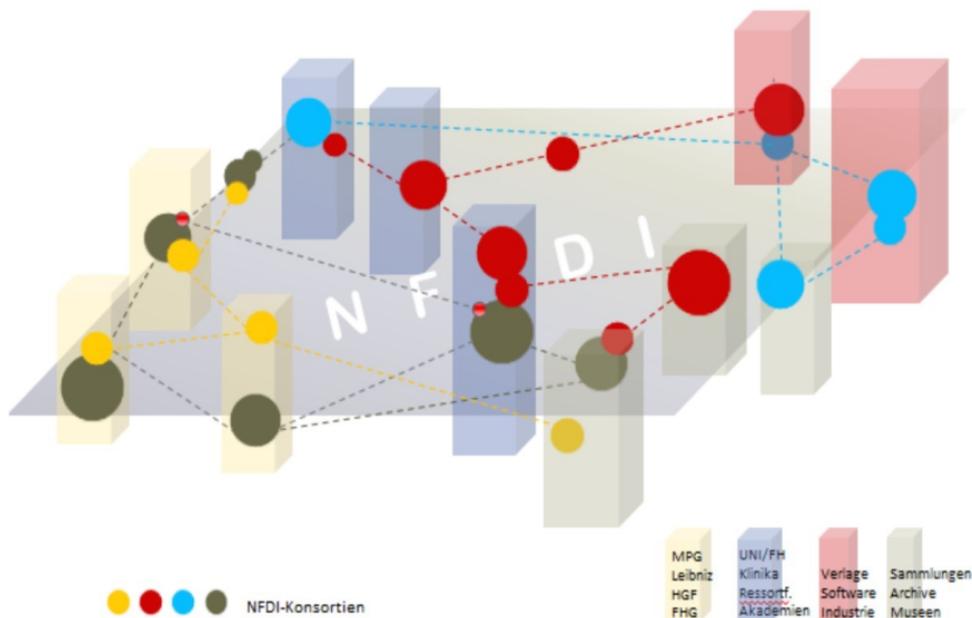


Abbildung 1.: Die Nationale Forschungsdateninfrastruktur als Querschnittsstruktur. Die NFDI besteht aus Konsortien, die sich aus Akteuren und Institutionen außeruniversitärer und Hochschulforschung, Industrie und weiteren Bereichen bilden (Quelle: Wambsganß, 2019, Folie 21).

Cloud⁹ spielen die fachlichen Communities für den Aufbau der Nationalen Forschungsdateninfrastruktur eine zentrale Rolle: Die Vernetzung erfolgt thematisch und wird daher stark von den in den Fachcommunities diskutierten und entwickelten Standards und Curricula beeinflusst.

Abseits des wissenschaftspolitischen Prozesses rund um die NFDI haben einige Bundesländer landesweite Aktivitäten zu Forschungsdatenmanagement (FDM) zu verzeichnen (Grasse/López/Winter 2018). Im Land Nordrhein-Westfalen werden die entsprechenden Aktivitäten in der Landesinitiative NFDI¹⁰ gebündelt. Diese nimmt eine Scharnierfunktion zwischen Landes- und Bundesaktivitäten wahr. Die wesentlichen Aufgaben der Landesinitiative sind die Vernetzung relevanter Stakeholder an Hochschulen und außeruniversitären Forschungseinrichtungen in NRW, sowie die Begleitung und Initiierung von hochschulübergreifenden Aktivitäten im Kontext von FDM, sodass mittel- bis langfristige NRW-weite Empfehlungen und Lösungen für FDM etabliert und darüber hinaus in einer nationalen Gesamtstrategie erfolgreich eingebracht und platziert werden können (Curdt, 2018). Als Ansprechpartnerin für Hochschulleitungen und Infrastrukturpartner im Bereich FDM, betreibt die Landesinitiative Trendscouting und sorgt für nationale sowie internationale Vernetzung. Aktuell befindet sich die Phase II für den Zeitraum 2019 bis 2021 in Planung und wird einen speziellen Fokus auf die Weiterbildung im Bereich E-Science und Forschungsdaten legen.

⁹ <https://www.egi.eu/about/newsletters/what-is-the-european-open-science-cloud/>

¹⁰ Landesinitiative NFDI der Digitalen Hochschule NRW. <https://fdm-nrw.de/>, besucht: 26.04.2019.

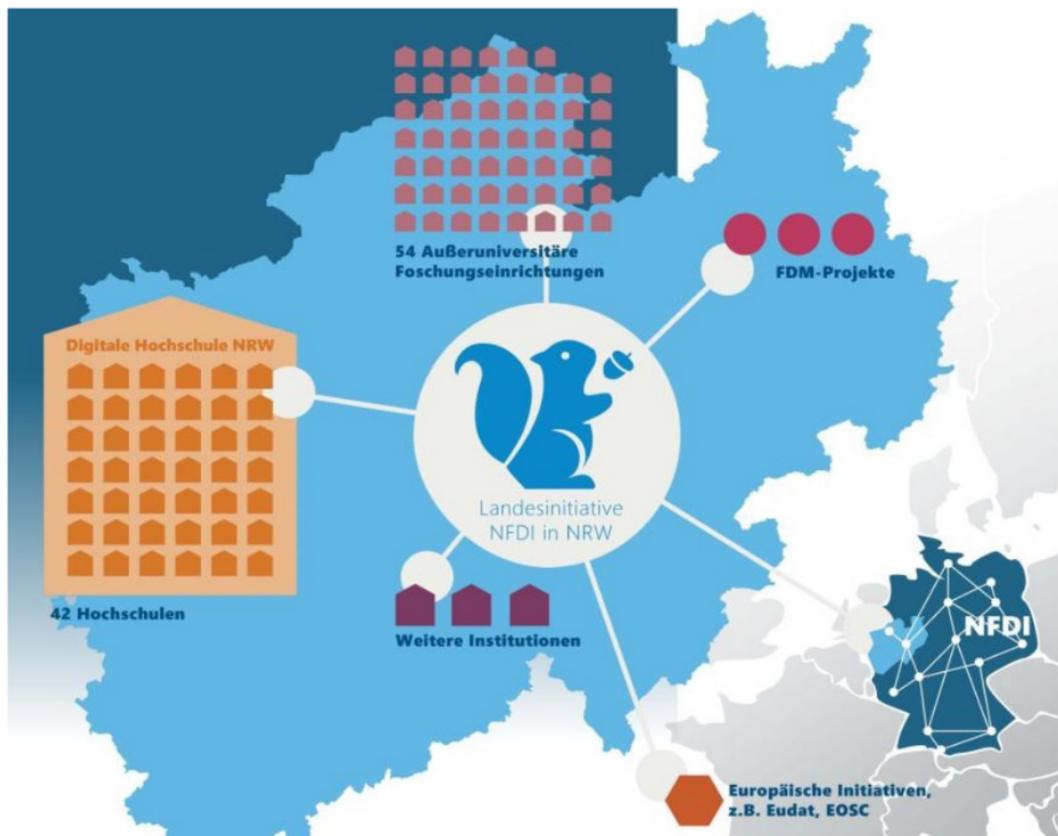


Abbildung 2.: Die Landesinitiative Nationale Forschungsdateninfrastruktur der Digitalen Hochschule NRW bietet Unterstützungsangebote für die Hochschulen und Forschungseinrichtungen in NRW im Kontext von Forschungsdatenmanagement und Fragen zur NFDI.

5. Überlegungen zur Rolle der Universitäten bei der Förderung und Weiterentwicklung von E-Science in den Disziplinen

Für die Universitäten stellt sich die Frage, ob sie die Bereitstellung entsprechender Angebote vor allem als eine Aufgabe von Fachgesellschaften und der Scientific Community bzw. einer nationalen Forschungsdateninfrastruktur ansehen oder ob sie die Dissemination und Weiterentwicklung von E-Science als Institution aktiv mitgestalten möchten. Hier besteht ein enger Zusammenhang mit der Schaffung von Rahmenbedingungen für wissenschaftliche Innovationen.

Bei den oben genannten sechs Perspektiven wurde am Beispiel der DSR schon deutlich, dass der Zugang zu bereits bestehenden Angeboten nicht ohne weiteres gegeben ist. Dies ist weitgehend übertragbar auf Forschungsdatenmanagement in anderen Fachdisziplinen. Die einzelne Wissenschaftlerin steht, gerade am Anfang ihrer Karriere, vor der Aufgabe, sich mit ihren spezifischen Forschungsinteressen und Karrierezielen in den in Abschnitt

2 skizzierten Perspektiven zu verorten und die Rolle von E-Science-Angeboten wie DSR im Rahmen der eigenen Aufgaben in Forschung und Lehre zu bestimmen. Die Universität Duisburg-Essen – als Beispiel – hält dazu bereits eine Reihe von Unterstützungsangeboten in den zentralen Einrichtungen Bibliothek (UB), Rechenzentrum („Zentrum für Informations- und Mediendienste“ (ZIM)) und Forschungsförderung („Science Support Center“ (SSC)) bereit.

Eine zentrale Rolle sieht die Universität dabei derzeit in Unterstützungsangeboten im Bereich Forschungsdatenmanagement (FDM) und baut hierzu die Servicestelle „Research Data Services“ (RDS) auf (siehe Abbildung 3). Das Konzept betont die zentrale Bedeutung von wechselseitigem Erfahrungsaustausch und Wissensaufbau zwischen den Wissenschaftler*innen der UDE und den Infrastruktureinrichtungen, um gezielt Unterstützungsangebote weiterzuentwickeln und mit bestehenden Initiativen der Fachcommunity zu verknüpfen. Im Rahmen eines kollegialen Kompetenznetzwerks soll der Austausch zwischen Wissenschaftler*innen untereinander sowie mit den Research Data Services dauerhaft etabliert werden.

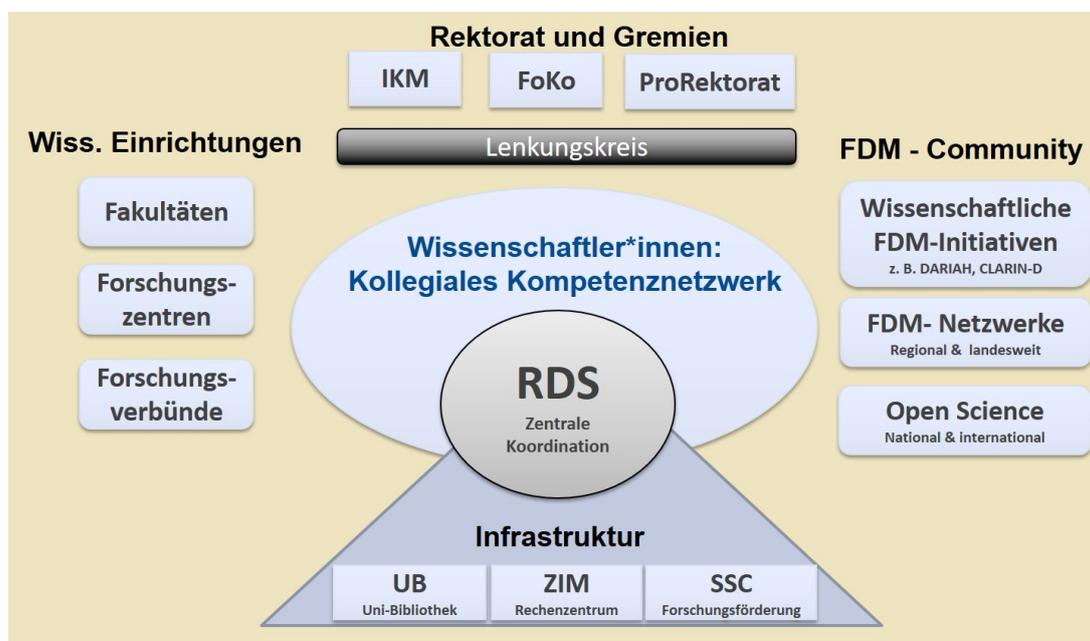


Abbildung 3.: Konzept eines Unterstützungsangebots „Research Data Services“ (RDS) an der Universität Duisburg-Essen (in Planung).

An dieser Stelle sei darauf hingewiesen, dass der Begriff E-Science in verschiedenen Fächern mit unterschiedlicher Schwerpunktsetzung verwendet wird und mit unterschiedlichen Anforderungen verknüpft ist. So ist z.B. das High Performance Computing (HPC) ein Bereich von E-Science, der in der Germanistischen Linguistik bislang noch keine prominente Rolle spielt, der in Bereichen insbesondere in den Natur- und Ingenieurwissenschaften hingegen von zentraler Bedeutung ist. Dort wo der Messaufbau und das analysierte Ergebnis für den Forschungsprozess meist wichtiger als die Rohdaten selbst sind, werden diese i.d.R. auch nicht aufbewahrt.

Im Gegensatz dazu sind in den Geisteswissenschaften oft die Rohdaten singulär oder originär in dem Sinne, dass sie durch Wiederholung nicht oder nur mit hohem Aufwand reproduziert werden könnten und dass einmal erhobene und mit sprach- und korpus-technologischen Verfahren aufbereitete Daten einen großen Wert darstellen, weil sie die Grundlage für anspruchsvolle empirische Forschung – im Falle von Referenzkorpora sogar für ganz unterschiedliche Forschungsfragen in einer Vielzahl von Projekten – bilden können. In Bezug auf die Germanistische Linguistik lassen sich aus den oben skizzierten Anforderungsbereichen Unterstützungsnahmen ableiten, die die Universität im Rahmen der RDS bieten bzw. anbieten könnte.

1. eine Lotsenfunktion übernehmen, die Forschende (insbesondere auch den wissenschaftlichen Nachwuchs) an für ihre Forschungsvorhaben geeignete Forschungsdateninfrastrukturen heranführt und sie mit entsprechenden Ansprechpartner*innen, Netzwerken und Zentren in Kontakt bringt;
2. die Verankerung von DSR in der Lehre unterstützen (unter anderem die Dokumentation von „Good Practices“, die für die weitere Dissemination genutzt werden können);
3. die Einbindung entsprechender Trainings- und Weiterbildungsmaßnahmen in universitäre Strukturen zur Förderung des wissenschaftlichen Nachwuchses wie Graduiertenkollegs, Graduate Centers initiieren und fördern;
4. Beratung in Bezug auf Forschungsinfrastrukturen und digitale Ressourcen – möglichst nah zu den Anforderungen im jeweiligen Fach bzw. Fächercluster – bei der Vorbereitung von Drittmittelanträgen anbieten;
5. technische und konzeptionelle Unterstützung für Forscher*innen bieten, die sich als Teil ihrer Forschungsarbeit am Auf- und Ausbau von digitalen Forschungsressourcen und -infrastrukturen beteiligen; und
6. das Engagement von Wissenschaftler*innen in Netzwerken und Fachgesellschaften unterstützen, die sich außeruniversitär mit der Weiterentwicklung von E-Science (Infrastruktur, Standards, Good Practices, Curricula) beschäftigen.

Dabei werden selbstverständlich vorhandene zentrale Infrastruktureinrichtungen, insbesondere die Bibliothek, das Rechenzentrum (ZIM) und die Forschungsförderung (SSC) eine wichtige Rolle spielen und durchgängig einzubinden sein. Gleichzeitig gilt: die Aufgaben beim Forschungsdatenmanagement gehen über das traditionelle Aufgabengebiet der Bibliotheken und Rechenzentren hinaus. Das gilt ganz besonders für die *Digital Humanities*, bei denen das *Computing* wie oben erläutert nicht im Mittelpunkt steht (Hinrichs, 2018). Wie bereits eingangs postuliert, sehen wir die Universitäten vor der Aufgabe, ihre Strategie zur E-Science in den Disziplinen weiterzuentwickeln – das Beispiel der Germanistische Linguistik zeigt, wie sehr die Umsetzung der eng verwobenen Anforderungen an Wissenschaftler*innen und Infrastruktureinrichtungen auf ein etabliertes Netzwerk mit gemeinsamen Kompetenzaufbau angewiesen ist.

6. Zusammenfassung und Ausblick

Die Möglichkeiten der Arbeit mit digitalen Daten, Ressourcen und Infrastrukturen verändern einerseits Forschungsprozesse und führen andererseits zu neuen Anforderungen in Bezug auf die Planung und Abwicklung von Forschungsprojekten, auf den Umgang mit und die Pflege von Forschungsdaten sowie auf die Vermittlung der für eine zeitgemäße, digital gestützte Forschung benötigten Kompetenzen an Studierende und Nachwuchswissenschaftler*innen. Am Beispiel der Germanistischen Linguistik haben wir den Stand der Arbeit mit digitalen Ressourcen und Infrastrukturen dargestellt und daraus eine Reihe von Anforderungen abgeleitet, die sich Forschenden und Lehrenden stellen, wenn sie entsprechende Ressourcen und die darauf bezogenen Methoden nachhaltig in Forschung und Lehre einsetzen möchten. Die Germanistische Linguistik diene uns dabei als eine Beispiel-Disziplin; die skizzierten Anforderungen lassen sich mit den erforderlichen Anpassungen sicherlich auf andere geisteswissenschaftliche Fächer übertragen. In anderen Disziplinen, etwa in den MINT-Fächern und im Bereich Medizin, mögen sich die Anforderungen in Bezug auf E-Science anders darstellen; auch hier gibt es aber Bedarfe, die für eine nachhaltige Verankerung der Arbeit mit digitalen Ressourcen und Infrastrukturen stellen.

Die Universitäten sind gefordert, ihr Selbstverständnis und ihre Rolle in Bezug auf die Förderung und Weiterentwicklung von E-Science in den Disziplinen zu entwickeln – im Sinne einer Bereitstellung von Rahmenbedingungen für wissenschaftliche Innovationen. Am Beispiel der Universität Duisburg-Essen haben wir skizziert, welche Unterstützungsangebote bereits existieren und wie diese weiterentwickelt werden können. Sehr sinnvoll erscheint es uns, den Prozess der Weiterentwicklung entsprechender Unterstützungsangebote unter Einbezug von Akteur*innen aus den Disziplinen bzw. Fakultäten zu gestalten („user Involvement“), um zwischen dem institutionell Realisierbaren und dem aus Sicht von Forschung Lehre Wünschenswerten einen guten und passgenauen Kompromiss zu finden und dabei geeignete Schnittstellen zu außeruniversitär existierenden Infrastrukturen, Services und Initiativen zu definieren. Die Förderung von E-Science und der nachhaltige Auf- und Ausbau von Forschungsdateninfrastrukturen ist eine grundsätzlich überuniversitär – auf nationaler und internationaler Ebene – zu leistende Aufgabe der Forschungsförderung (vgl. z. B. Hinrichs 2018). Die Anbindung universitärer Forschung und Lehre an diese Infrastrukturen und die Verankerung ihrer Nutzung in den Disziplinen kann aber sinnvoll nur auf Ebene der Universitäten geleistet werden. Dazu sind einerseits technische und konzeptionelle Unterstützungsangebote erforderlich, die möglichst nahe an den Bedarfen einzelner Disziplinen und Disziplinen-Cluster ansetzen, andererseits Anreizsysteme, um die Dissemination der Potenziale digitaler Forschungsinfrastrukturen und deren Nutzung in den Disziplinen nachhaltig zu verankern.

Literaturverzeichnis

- [1] Beißwenger, Michael (2018): Internetbasierte Kommunikation und Korpuslinguistik: Repräsentation basaler Interaktionsformate in TEI. In: Henning Lobin, Roman Schneider and Andreas Witt (Hrsg.): Digitale Infrastrukturen für die germanistische Forschung. Berlin/New York: de Gruyter 2018 (Germanistische Sprachwissenschaft um 2020, Bd. 6), 307-349. Open-Access-Publikation: <https://doi.org/10.1515/9783110538663-015>
- [2] Beißwenger, Michael; Chanier, Thierry; Erjavec, Tomaž; Fišer, Darja; Herold, Axel; Lubešic, Nikola; Lungen, Harald; Poudat, Céline; Stemle, Egon; Storrer, Angelika; Wigham, Ciara (2017): Closing a Gap in the Language Resources Landscape: Groundwork and Best Practices from Projects on Computer-mediated Communication in four European Countries. In: Lars Borin (Ed.): Selected papers from the CLARIN Annual Conference 2016, Aix-en-Provence, 26–28 October 2016, CLARIN Common Language Resources and Technology Infrastructure (Linköping University Electronic Conference Proceedings 136), 1-18.
- [3] Curdt, Constanze, Grasse, Marleen, Hess, Volker, Kasties, Nils, López, Ania, Magrean, Benedikt, Winter, Nina. (2018). ZUR ROLLE DER HOCHSCHULEN - Positionspapier der Landesinitiative NFDI und Expertengruppe FDM der Digitalen Hochschule NRW zum Aufbau einer Nationalen Forschungsdateninfrastruktur. Zenodo. <http://doi.org/10.5281/zenodo.1217527> DFG (2015a): Handreichung: Empfehlungen zu datentechnischen Standards und Tools bei der Erhebung von Sprachkorpora. www.dfg.de/download/pdf/foerderung/grundlagen_dfg_foerderung/informationen_fachwissenschaften/geisteswissenschaften/standards_sprachkorpora.pdf
- [4] DFG (2015b): Handreichung: Informationen zu rechtlichen Aspekten bei der Handhabung von Sprachkorpora. https://www.dfg.de/download/pdf/foerderung/grundlagen_dfg_foerderung/informationen_fachwissenschaften/geisteswissenschaften/standards_recht.pdf
- [5] Grasse, M., López, A. and Winter, N., (2018). Landesinitiative NFDI – a Central Point of Contact for RDM for Higher Education Institutions in the German State of North Rhine-Westphalia. Data Science Journal, 17, p.25.DOI:<http://doi.org/10.5334/dsj-2018-025>
- [6] Geyken, Alexander, Adrien Barbaresi, Jörg Didakowski, Bryan Jurish, Frank Wiegand & Lothar Lemnitzer (2017): Die Korpusplattform des „Digitalen Wörterbuchs der deutschen Sprache“ (DWDS). In: Zeitschrift für germanistische Linguistik 45 (2), 327–344.
- [7] Hinrichs, Erhard (2018): Digitale Forschungsinfrastrukturen für die Sprachwissenschaft. In: Lobin, H. (Hrsg.): Digitale Infrastrukturen für die germanistische Forschung, S. 81–133. De Gruyter, Berlin.

- [8] Lemnitzer, Lothar & Heike Zinsmeister (2015): *Korpuslinguistik. Eine Einführung*. 3., überarbeitet und revidierte Auflage. Tübingen: Narr (Narr Studienbücher).
- [9] Lüdeling, Anke; Kytö, Merja (2008/2009): *Corpus Linguistics*. 2 Bde. Berlin/New York: de Gruyter (Handbücher zur Sprach- und Kommunikationswissenschaft 29).
- [10] Lungen, Harald (2017): DeReKo – Das Deutsche Referenzkorpus. Schriftkorpora der deutschen Gegenwartssprache am Institut für Deutsche Sprache in Mannheim. In: *Zeitschrift für germanistische Linguistik* 45 (1), 161–170.
- [11] Schmidt, Thomas (2017): DGD – die Datenbank für Gesprochenes Deutsch. Mündliche Korpora am Institut für Deutsche Sprache (IDS) in Mannheim. In: *Zeitschrift für germanistische Linguistik* 45 (3), 451-463.
- [12] Wambsganß, Joachim (2019): Rat für Informationsinfrastrukturen (RfII) - Einblicke in aktuelle Themen. und Arbeiten. E-Science-Tage Heidelberg, Keynote.