

# Pragmatisches Forschungsdatenmanagement Qualitative und quantitative Analyse der Bedarfslandschaft in den Computational Literary Studies

## Helling, Patrick

patrick.helling@uni-koeln.de  
Institut für Deutsche Philologie, Lehrstuhl für  
Computerphilologie und neuere deutsche Literaturgeschichte,  
Universität Würzburg

## Jung, Kerstin

kerstin.jung@ims.uni-stuttgart.de  
Institut für Deutsche Philologie, Lehrstuhl für  
Computerphilologie und neuere deutsche Literaturgeschichte,  
Universität Würzburg

## Pielström, Steffen

pielstroem@biozentrum.uni-wuerzburg.de  
Institut für Deutsche Philologie, Lehrstuhl für  
Computerphilologie und neuere deutsche Literaturgeschichte,  
Universität Würzburg

## Einleitung

Die *Computational Literary Studies* (CLS) sind ein aufstrebendes, interdisziplinäres Forschungsfeld, in dem Gegenstände und Fragestellungen aus der Literaturwissenschaft mit computergestützten, teilweise quantitativen Methoden bearbeitet werden. Damit verorten sich die CLS am Schnittpunkt von Literaturwissenschaft, Computerlinguistik und Informatik.

Bedingt durch diese digitalen Methoden spielen Forschungsdaten unterschiedlichster Art eine zentrale Rolle für die CLS: Die Basis eines jeden Projektes stellt ein Korpus digitalisierter literarischer Texte dar. Dazu kommen weitere Arten von Forschungsdaten, unter anderem Textannotationen, zusammenfassende Statistiken und Visualisierungen, Metadaten und, bedingt durch den aktuellen Deep-Learning-Trend in der Computerlinguistik, zunehmend auch komplexe statistische Sprachmodelle. Entsprechend zeichnet sich die Forschungsdatenmanagement-Bedarfslandschaft der CLS durch eine starke Heterogenität aus, die auch in vielen anderen Teildisziplinen der Geisteswissenschaften festzustellen ist (Pempe 2012). Der Umgang mit diesen Forschungsdaten über den gesamten Forschungsdatenlebenszyklus hinweg stellt dabei eine Grundbedingung wissenschaftlichen Fortschritts dar (Bryant, Lavoie & Maipas 2017) und ist nicht erst seit Bestrebungen hin zu einer fachlich getriebenen Nationalen Forschungsdateninfrastruktur (RfII 2016, 2017) wesentlicher Bestandteil guter wissenschaftlicher Praxis (DFG 2019).<sup>1</sup> Für die Transparenz der wissenschaftlichen Methode und die Reproduzierbarkeit der Ergebnisse ist ein fachspezifisches Management der zu Grunde liegenden Forschungsdaten im Sinne der FAIR-Prinzipien (Wilkinson et al. 2016) bis hin zur nachhaltigen

Publikation und Archivierung auch in den CLS von zentraler Bedeutung. Einige spezifische Aspekte, wie die Verwendung nicht exakt reproduzierbarer, stochastischer Verfahren oder die oft komplexe rechtliche Situation der teilweise urheberrechtlich geschützten Primärdaten stellen hierbei besondere Herausforderungen dar (vgl. Schöch et al. 2020; Kleinkopf et al. 2021).

Das DFG Schwerpunktprogramm SPP 2207 „Computational Literary Studies“ (SPP CLS) setzt sich aus insgesamt 11 an verschiedenen Universitäten in Deutschland und der Schweiz angesiedelten Forschungsprojekten und einem Datenkoordinationsteam zusammen.<sup>2</sup>

Das Datenkoordinationsteam ist mit zwei halben Stellen sowie einer Koordinierungsstelle ausgestattet und zentral an der Gesamtkoordinationsstelle des SPP CLS angesiedelt. Seine Mitglieder verfügen sowohl über langjährige Erfahrungen und Kompetenzen im methodischen Bereich der Computational Literary Studies als auch im fachspezifischen, geisteswissenschaftlichen Forschungsdatenmanagement (FDM). Zusätzlich zur Entwicklung und Umsetzung einer gemeinsamen Strategie für das Management von Forschungsdaten für das gesamte SPP unterstützt das Team auch die Koordination des gesamten Schwerpunktprogramms.

Das SPP CLS bietet durch seine Bündelung von verschiedenen CLS-Forschungsvorhaben einen hervorragenden Rahmen, um zu untersuchen, welche Art von Forschungsdaten in CLS-Projekten wie genutzt werden. Ziel einer solchen Landschaftsvermessung ist es, die Forschungspraxis im SPP 2207 zu erfassen und zu beobachten, um daraus *Best Practices* im Umgang mit Forschungsdaten zur Schaffung eines Mehrwerts für das gesamte Feld zu identifizieren und zu aggregieren sowie methodisch verwandte Fachbereiche wie bspw. textbasiert arbeitende Digital Humanities oder die Computerlinguistik mit zu adressieren.

Darüber hinaus können die methodischen Ansätze zur Entwicklung einer Strategie für das Forschungsdatenmanagement sowie die strukturelle und organisatorische Einbindung des Datenkoordinationsteams in den Gesamtkontext des Schwerpunktprogramms als ein konkret erprobtes Praxisbeispiel für die Bedienung von FDM-Bedarfen und das Management von Forschungsdaten innerhalb von Forschungs- und Infrastrukturverbänden wie bspw. Sonderforschungsbereichen, Exzellenzclustern oder grundsätzlich übergreifenden Informationsinfrastrukturprojekten verstanden werden.

## Vorgehen zur Landschaftsvermessung

Zur Analyse der Bedarfe zum Forschungsdatenmanagement in den CLS wurde ein Vorgehen mit Interviewgesprächen sowie mehreren Analyseschritten und Reviewphasen entwickelt. Durch diesen Doppelschritt konnte sich bei der Datenerfassung besonders nah am tatsächlichen Forschungsalltag sowie den aktuellen Bedingungen und Bedarfen jedes einzelnen Projekts orientiert werden. Dies kann bei bspw. quantitativen, ggf. sogar anonym durchgeführten, Onlineumfragen zu FDM-Bedarfen, die durchaus fehleranfällig sein können, nicht zwangsläufig gewährleistet werden, da hier i.d.R. keine Möglichkeit besteht auf die Antworten der Befragten konkreter einzugehen.

Zunächst wurde mit jedem Projekt ein Interview auf Basis eines Leitfadens aus 47 offenen, nach Projektphasen gruppierten Fragen durchgeführt<sup>3</sup>: (i) zum Umgang mit Daten und lebenden Systemen im laufenden Projekt, sowie (ii) zu Publikations- und Archivierungsstrategien am Ende des Projekts. Dabei dienten die qualitativen Interviews neben der Landschaftsvermessung auch

dem Kennenlernen der jeweiligen Projekte sowie dem Aufbau der Kommunikation zwischen den Projektbeteiligten und dem Datenkoordinationsteam.

Im Interview wurden allgemeine Beispiele zur Erläuterung der Fragen angegeben. Aus den gegebenen Antworten wurde ein Antworteninventar erstellt, das den Projekten zusammen mit den eigenen Antworten zum Review zur Verfügung gestellt wurde. Dieser Schritt stellte sicher, dass die Antworten der Projekte korrekt zugeordnet wurden und, dass Aspekte, die auf mehrere Projekte zutreffen, aber nicht von allen erwähnt wurden, am Ende dennoch für die folgenden Analysen umfassend erfasst werden konnten.

Dabei wurden Review und Analyse zunächst auf Aspekte zum laufenden Projekt konzentriert, da Fragen zum Ende der Projektphase eher tentativ beantwortet wurden. Zum Zeitpunkt dieser Einreichung befanden sich die Fragen zum Ende der Projektphase im Reviewprozess.

Für die Landschaftsvermessung der CLS in Bezug auf (1) wissenschaftliches Arbeiten, (2) Management von Forschungsdaten sowie entsprechende (3) Trends, (4) Best Practices und (5) community-getriebene Standards sowie die Entwicklung einer gemeinsamen Datenstrategie wurden die Interviews zunächst quantitativ ausgewertet (siehe Abschnitt 3). Zur Umsetzung pragmatischer Lösungsstrategien für das gesamte SPP CLS durch die Datenkoordination im Sinne der Identifikation und Umsetzung von FDM-Lösungsstrategien orientiert an der (a) Gesamtheit der Bedarfe im Schwerpunktprogramm und auf der (b) Basis existierender Werkzeuge und Angebotsstrukturen in der gesamten FDM-Landschaft, war neben dieser quantitativen Analyse auch eine qualitative Auswertung der Inhalte relevant (siehe Abschnitt 4).

## Quantitative Auswertung: Erste Ergebnisse der Landschaftsvermessung

Zur Beschreibung der Datenlandschaft und Entwicklung einer passgenauen Datenstrategie für das gesamte SPP CLS ist zentral, welche Datentypen und -formate genutzt und produziert werden. Im Kontext der digitalen Literaturwissenschaften vermeintlich wenig überraschend arbeiten alle Projekte des Schwerpunktprogramms mit Textdaten und beinahe genauso viele mit Softwarecode. Aber auch numerische und bibliographische Daten, sogar Bilddaten spielen bei einigen Projekten eine wichtige Rolle (Tabelle 1) und müssen bei der Archivierung und Nachnutzbarkeit von Projektergebnissen mitberücksichtigt werden.

Tabelle 1: Genutzte Datentypen im SPP CLS.

Datentypen	Projekte
Text	10
Softwarecode	9
Numerische Daten	6
Bilddaten	5
Bibliographische Daten	4
Wörterbücher/Listen	2
Interviewdaten	2
Netzwerkdaten	1

Tab. 1: Genutzte Datentypen im SPP CLS.

Mit XML, PlainText-Formaten sowie PDF nutzen viele Projekte textbasierte Datenformate, die sich bereits vergleichsweise gut für eine nachhaltige Archivierung und Nachnutzung eignen. Ähnliches gilt für die Nutzung von CSV-Dateien. Dennoch wird deutlich, dass in der Bandbreite genutzter Formate einige Projekte auch proprietäre Lösungen verwenden (Abb. 1), wodurch aus FDM-Perspektive gegen Projektende eine Formatmigration nötig werden könnte.

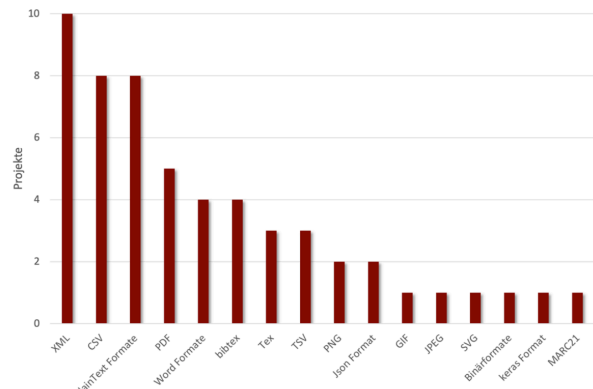


Abb. 1: Genutzte Datenformate im SPP CLS.

Vor dem Hintergrund der hohen Relevanz von Softwarecode (vgl. Tabelle 1) spielt auch die Nutzung von Programmier- und Skriptsprachen in einem zentralen Datenmanagement eine wichtige Rolle (Tabelle 2).

Programmier-/Skriptsprache	Projekte
Python	9
R	4
Shell-Skripte	4
Java	3
X-Technologien	2
JavaScript	2
HTML	2
CSS	2
SQL	1

Tab. 2: Genutzte Programmier- und Skriptsprachen im SPP CLS.

In diesem Zusammenhang sind gleichzeitig der Umgang mit lebenden Systemen (Tabelle 3) und hier verwendeter Technologie-Stacks (Abb. 2) am Ende der Projektlaufzeit eine zentrale Herausforderung. Insbesondere die durch die Projekte teilweise selbst motivierte Nutzung von statischen Systemen wie bspw. Jekyll wird die langfristige Verfügbarkeit von lebenden Systemen dabei deutlich erleichtert.

Lebende Systeme	Projekte
Website	7
Tools/Anwendungen	3
Bibliotheken	1
Dashboard	1

Tab. 3: Geplante lebende Systeme im SPP CLS.

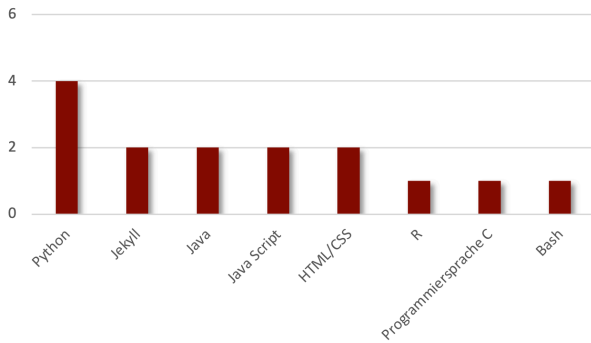


Abb. 2: Geplante Technologie-Stacks zur Entwicklung lebender Systeme im SPP CLS.

Neben der Erfassung von Informationen, die für das FDM relevant sind, war es bei der Landschaftsvermessung im SPP CLS auch ein Ziel Aussagen über methodische und organisatorische Best Practices innerhalb des Forschungsfelds zu treffen.

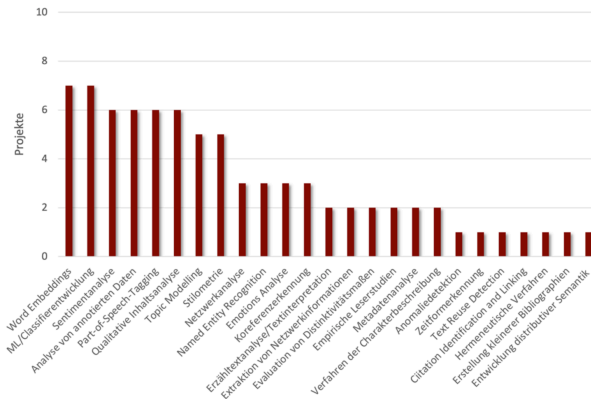


Abb. 3: Analyseverfahren und methodische Werkzeuge im SPP CLS.

Während die Abfrage von angewandten Analyseverfahren und methodischer Werkzeuge innerhalb der einzelnen Projekte erste Trends ablesen lassen (Abb. 3), können Informationen über genutzte Tools zum Projektmanagement sowie zur Annotation von Daten dabei helfen, infrastrukturelle Bedarfe der Community zu identifizieren (Abb. 4 und Tabelle 4).

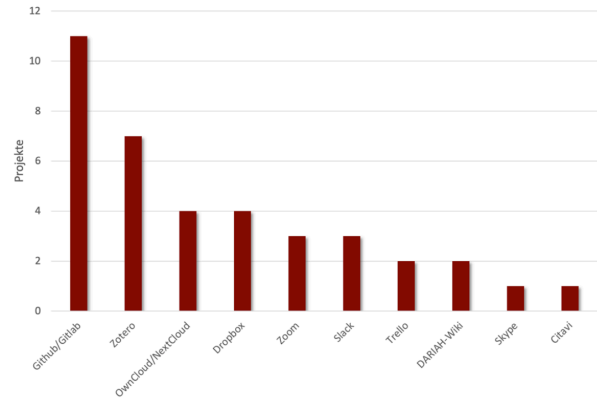


Abb. 4: Genutzte Tools zur Projektverwaltung im SPP CLS.

Annotationstool	Projekte
Catma	5
TreeTagger	1
CorefAnnotator	1
Sentiment Analyzer	1

Tab. 4: Genutzte Annotationstools im SPP CLS.

Darüber hinaus können die Angaben zu Archivierungs- und Publikationsstrategien, trotz ihres noch tentativen Charakters, bereits für die Ableitung von Best Practices innerhalb der Fachdisziplin genutzt werden (Abb. 5 und Abb. 6).

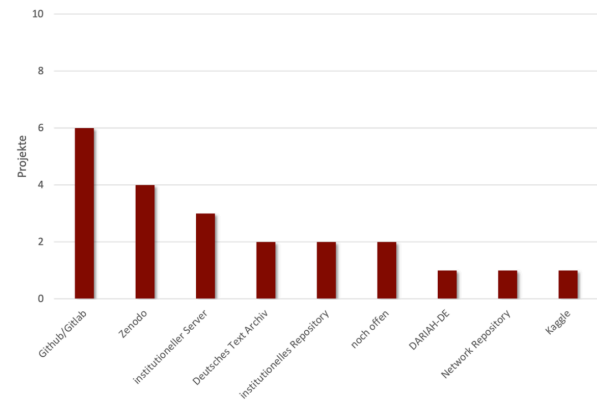


Abb. 5: Tentative Angaben zu genutzten Archivierungsinfrastrukturen im SPP CLS.

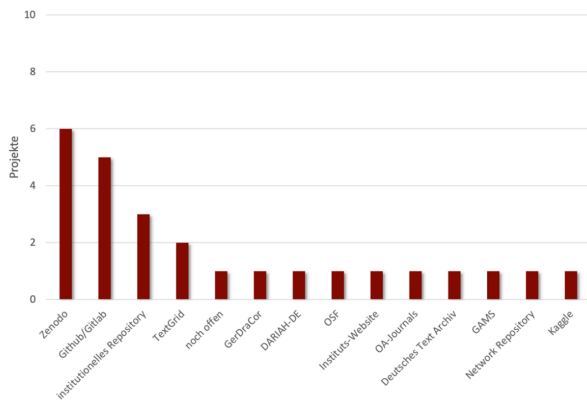


Abb. 6: Tentative Angaben zu genutzten Publikationsinfrastrukturen im SPP CLS.

## Qualitative Auswertung: Abgeleitete FDM-Maßnahmen

Beispielhaft für die pragmatische Identifikation und Umsetzung von Maßnahmen zur Handhabung der Bedarfe im Forschungsdatenmanagement in den CLS gehen wir hier auf zwei Aspekte ein.

Zum Arbeiten in und zum Austausch zwischen den Projekten sowie der projektübergreifenden Arbeit einzelner Arbeitsgruppen wurde nach einer geeigneten Plattform gesucht. Dabei waren folgende Kriterien ausschlaggebend:

- Die Zusammenarbeit muss über Fach-, Universitäts-, und Ländergrenzen vollumfänglich möglich sein.
- Die Ablage und der Austausch von Daten muss möglich sein, idealerweise auch das gemeinsame Arbeiten auf der Plattform, die daher zumindest eine Versionierung zur Verfügung stellen muss.
- Textdateien und Annotationen müssen ebenso verwaltet werden können wie formale Metadaten und Softwarecode.
- Die Gesamtkapazität muss im hohen GB-Bereich liegen, da Datensätze sowie Modelle aus dem Maschinellen Lernen im zweistelligen GB-Bereich als Einzeldateien zu erwarten sind.
- Bereiche für Projekte, Projektgruppen sowie Arbeitsgruppen müssen leicht angelegt und verwaltet werden können.
- Der Speicherort muss bekannt und sicher sein, damit rechtliche Belange, bspw. urheberrechtlicher Natur, gewahrt und dem Missbrauch der Daten vorgebeugt werden kann.

Cloud-Lösungen sind für den Datenaustausch und das kollaborative Arbeiten oft die erste Wahl. Jedoch sind sie meist auf bestimmte, ggf. lokale Nutzergruppen ausgerichtet (z.B. Dienste für Hochschulen eines Bundeslandes), bezüglich des Speicherorts intransparent oder mit kommerziell tätigen Unternehmen verbunden, was durch unterschiedliche Richtlinien der Universitäten ebenfalls nicht alle beteiligten Projekte einschließen kann.

Letzteres trifft auch auf zentral zugreifbare Entwicklungsplattformen wie Github oder Gitlab zu, die allerdings den zusätzlichen Vorteil haben, dass sie auch für gemeinsame Codeentwicklung zur Verfügung stehen und Features zum Projektmanagement, wie z.B. Ticketsysteme anbieten.

Wengleich es einen wachsenden, disziplinübergreifenden Bedarf an der Nutzung von kollaborativen Versionierungssystemen gibt, der sich u.a. aus den mittlerweile breit aufgestellten Schu-

lungs- und Workshop-Angeboten ableiten lässt,<sup>4</sup> gibt es gleichzeitig einen Mangel an standortübergreifend nutzbaren, zentralen Angeboten solcher Systeme. Entsprechend wurde, trotz eines hohen Betreuungs- und Verwaltungsaufwandes, eine eigene Gitlab-Instanz für das SPP CLS auf universitären Servern aufgesetzt. Dabei war neben der Erfüllung der Kriterien ausschlaggebend, dass im Schwerpunktprogramm bereits Erfahrungen im Umgang mit git-basierten Lösungen vorhanden waren. Für große Einzeldateien wurde das sogenannte Large File Storage zur Verfügung gestellt. Backups erfolgen durch die Infrastruktur des universitären Rechenzentrums. Die Instanz wird durch das Datenkoordinationsteam des SPP CLS betrieben und verwaltet. Da durch die große Funktionalität von Gitlab eine gezielte Verwendung mit Einstiegshürden verbunden sein kann, wurde eine spezifische Dokumentation zusammengestellt und ein Einstiegsworkshop für die Mitglieder des Schwerpunktprogramms organisiert.

Ein zweiter Aspekt ist der Bedarf einer zentralen Publikationsplattform, z.B. für Materialien, die nicht in fachspezifischen Repositorien oder Publikationsorganen unterkommen (Posterpräsentationen, Folien, Handreichungen, aber perspektivisch auch Datendumps oder Snapshots lebender Systeme). Wichtige Anforderungen sind hierbei eine langfristige Auffindbarkeit und Zitierbarkeit sowie die Möglichkeit Ergebnisse der verschiedenen Projekte im Projektverbund gemeinsam sichtbar zu machen:

- Die abgelegten Daten sollen bei einer Speicherinstitution liegen, bei der klar ist, wo die entsprechende Infrastruktur unterhalten wird und wer darauf Zugriff hat.
- Die Dauerhaftigkeit der Speicherinstitution sollte gegeben sein.
- Die Ablage von Daten sollte, auch in größeren Mengen, für Forschende ohne zusätzliche Kosten möglich sein.
- Eine maximale Daten-/Dateiobergrenze sollte es nicht geben. Der verfügbare Speicherplatz sollte mindestens im zweistelligen GB-Bereich liegen.
- Die Speicherinstitution sollte die Vergabe von persistenten Identifiern ermöglichen.
- Mit Hilfe von Versionierung und möglichst auch versionierbarer, persistenter Identifier sollten verschiedene Zustände von Publikationen, Daten und sonstigen Materialien veröffentlicht werden können.
- Die Vergabe von Lizenzen und Möglichkeit eines abgestuften Zugriffs sollte unterstützt werden, um ggf. auch rechtlich geschützte Materialien gesammelt abzulegen.
- Technische Komponenten des Systems sollten transparent sein.
- Zur Steigerung der Auffindbarkeit und Nachnutzung von Publikationen sollte die Speicherinstitution Schnittstellen zu anderen Portalen anbieten und Metadaten an weitere Onlinekataloge weitergeben.

Unterschiedliche projektfinanzierte Repositorien und Publikationssysteme stellen zwar mögliche Lösungen für die skizzierten Anforderungen dar, allerdings können Förderstrategien, insbesondere innerhalb der deutschsprachigen Wissenschaftslandschaft, deren langfristige Weiterfinanzierung und somit die Dauerhaftigkeit eines Services i.d.R. nicht gewährleisten. Bereits institutionalisierte Angebotsstrukturen verfügen hingegen häufig entweder über eine Begrenzung des Adressatenkreises, oder haben sich fachlich oder formatspezifisch stark spezialisiert. Die Nutzung von wirtschaftlich-kommerziellen Angeboten kann wiederum, abgesehen von wissenschaftsethischen Einwänden und beschränkenden universitären Richtlinien, von den nutzenden Projekten und Wissenschaftler\*innen selbst auf Dauer nicht getragen und finanziert werden.

Zur Bedienung der skizzierten Bedarfe wurden zwei Lösungen identifiziert und in die Datenstrategie des SPP CLS integriert: Zunächst wurde eine eigene Community für das Schwerpunktprogramm auf dem Online-Speicherdienst Zenodo eingerichtet, welcher nahezu allen Anforderungen entspricht.<sup>5</sup> Das generische Repositorium ist mittlerweile in vielen Fachdisziplinen als Dienst etabliert. Es ermöglicht die Verwendung reichhaltiger Metadaten zur Beschreibung von Publikationen, unterstützt die Vergabe von versionierten Digital Object Identifiern (DOI), gibt Metadaten an aggregierende Portale wie bspw. OpenAIRE weiter und wird vom CERN in der Schweiz dauerhaft betrieben.<sup>6</sup>

Für die langfristige Veröffentlichung von lebenden Systemen wie bspw. Websites, Tools und einfacher Anwendungen stellt in Ergänzung Github eine Lösung dar. Auch der netzbasierte Dienst zur Versionsverwaltung verfügt in diversen wissenschaftlichen Communities, obwohl er mittlerweile von Microsoft betrieben wird, über einen großen Nutzendenkreis und kann für den abgegrenzten Gegenstandsbereich einiger lebender Systeme in Frage kommen: Github ermöglicht neben der kollaborativen Entwicklung auch die dokumentierte und quelloffene Bereitstellung von Software und verfügt über eine Schnittstelle zu Zenodo, wodurch Github-Repositorien in einem bestimmten Zustand auf Zenodo publiziert und persistent referenzierbar gemacht werden können.

## Weitere Perspektiven auf eine gemeinsamen Datenstrategie

Trotz erster pragmatischer und bedarfsorientierter Lösungsstrategien im Rahmen der Entwicklung einer gemeinsamen Datenstrategie innerhalb des SPP CLS hat die erste Review- und Analysephase der Landschaftsvermessung deutlich gemacht, dass innerhalb des Schwerpunktprogramms sehr heterogene Bedingungen und Bedarfe in Bezug auf das Forschungsdatenmanagement vorherrschen. Unterschiedliche Methoden angewandt auf verschiedene Korpora erzeugen teilweise individuelle Forschungsdaten und -ergebnisse, die es sowohl projektintern als auch für das gesamte SPP CLS langfristig zu sichern sowie verfügbar zu machen gilt.

Eine zentrale Herausforderung der zweiten Review- und Analysephase, mit Fokus auf der Konkretisierung individueller Archivierungs- und Publikationsstrategien innerhalb der einzelnen Projekte, ist die Integration etablierter Vorgehensweisen in die gemeinsame Datenstrategie des SPP CLS. Darüber hinaus wird es eine Hauptaufgabe sein die Entwicklung von lebenden Systemen innerhalb der einzelnen Projekte so weit zu begleiten und zu betreuen, dass möglichst alle individuellen Websites, Tools und kleineren Anwendungen auch über die Projektphasen hinaus in einer statischen Form mit geringem Kurationsaufwand, bspw. via Github und Zenodo, auffindbar, zugänglich, interoperabel und nachnutzbar bleiben.

In unserem Vortrag werden wir das Schwerpunktprogramm als Blaupause für das Forschungsdatenmanagement innerhalb der CLS dezidiert beschreiben und die hier beschriebenen Ergebnisse mit weiteren Erkenntnissen, auch zu möglichen fachspezifischeren Lösungen, die wir aus der bis dahin abgeschlossenen zweiten Review- und Analysephase gewinnen werden, kompletieren.

## Fußnoten

1. Vgl. <https://www.nfdi.de/> und <https://www.gwk-bonn.de/themen/weitere-arbeitsgebiete/informationsinfrastrukturen-nfdi/> (letzter Zugriff: 14. Juli 2021).
2. DFG Schwerpunktprogramm „Computational Literary Studies“, Online: <https://dfg-spp-cls.github.io/> (letzter Zugriff: 09. Juli 2021).
3. Interviewleitfaden zur FDM-Bestandsaufnahme im Schwerpunktprogramm „Computational Literary Studies“, Online: <http://doi.org/10.5281/zenodo.4269639>.
4. Vgl. u.a. diverse Workshops zum Forschungsdatenmanagement mit Gitlab, durchgeführt durch die Landesinitiative für Forschungsdatenmanagement (fdm.nrw), Online: <https://www.fdm.nrw/index.php/fdm-nrw/versionierung-gitlab/> (letzter Zugriff: 09.11.2021); „Workshop: Git und Gitlab für Anfänger\*innen“, durchgeführt durch die Landesinitiative FDM Thüringen, 21. Juli 2021, Online: <https://forschungsdaten-thueringen.de/veranstaltung/workshop-git-gitlab-de.html> (letzter Zugriff: 09.11.2021); „git and GitLab basics workshop“, durchgeführt durch NFDI4Ing, 21.09.2021, Online: <https://nfdi4ing.de/git-and-gitlab-basics-workshop-3/> (letzter Zugriff: 09.11.2021); Workshop „Datenversionierung mit Git – Advanced Track“, durchgeführt durch Carolin Odebrecht auf der RDA Deutschland Tagung 2020, Potsdam, 25.02.2020, Online: <https://www.rda-deutschland.de/presentationen-2020/gitlabworkshop2020.pdf> (Folien) (letzter Zugriff: 09.11.2021).
5. Online: <https://zenodo.org/>
6. Online: <https://www.openaire.eu/>

## Bibliographie

- Bryant, Rebecca / Lavoie, Brian / Malpas, Constance** (2017): *A Tour of the Research Data Management (RDM) Service Space. The Realities of Research Data Management, Part 1*. Dublin, Ohio: OCLC Research. DOI: <https://doi.org/10.25333/C3PG8J>.
- DFG - Deutsche Forschungsgemeinschaft** (2019): *Guidelines for Safeguarding Good Research Practice. Code of Conduct*. Zenodo: <http://doi.org/10.5281/zenodo.3923602>.
- Kleinkopf, Felicitas / Jacke, Janina / Gärtner, Markus** (2021): „Text- und Data-Mining: urheberrechtliche Grenzen der Nachnutzung wissenschaftlicher Korpora und ihrer Bedeutung für die Digital Humanities“ in: *MMR: Zeitschrift für IT-Recht und Recht der Digitalisierung*, Jahrgang 2021, Heft 3. München: C.H.BECK oHG 196 ff. Online: <http://dx.doi.org/10.18419/opus-11445>.
- Pempe, Wolfgang** (2012): „Geisteswissenschaften“ in: Neuroth, Heike / Strathmann, Stefan / Oßwald, Achim / Scheffel, Regine / Klump, Jens / Ludwig, Jens (eds.): *Langzeitarchivierung von Forschungsdaten. Eine Bestandsaufnahme*. Boizenburg: Verlag Werner Hülsbusch 137-160.
- RfII 2016, RfII - Rat für Informationsinfrastrukturen** (2016): *Leistung aus Vielfalt. Empfehlungen zu Strukturen, Prozessen und Finanzierung des Forschungsdatenmanagements in Deutschland*. Göttingen. Online: <https://d-nb.info/1104292440/34> (letzter Zugriff: 14. Juli 2021).
- RfII - Rat für Informationsinfrastrukturen** (2017): *Schritt für Schritt - oder: Was bringt wer mit? Ein Diskussionsimpuls für den Einstieg in die Nationale Forschungsdateninfrastruktur (NFDI)*. Göttingen, Online: <https://d-nb.info/1131083113/34> (letzter Zugriff: 14. Juli 2021).

Schöch, Christof / Döhl, Frédéric / Rettinger, Achim / Gius, Evelyn / Trilcke, Peer / Leinen, Peter / Jannidis, Fotis / Hinzmann, Maria / Röpke, Jörg (2020): „Abgeleitete Textformate: Text und Data Mining mit urheberrechtlich geschützten Textbeständen“ in: *Zeitschrift für digitale Geisteswissenschaften*. Wolfenbüttel, Online: [https://doi.org/10.17175/2020\\_006](https://doi.org/10.17175/2020_006).

Wilkinson, Mark D. / Dumontier, Michel / Aalbersberg, IJsbrand Jan / Appleton, Gabrielle / Axton, Myles / Baak, Arie / Blomberg, Niklas / Boiten, Jan-Willem / da Silva Santos, Luiz Bonino / Bourne, Philip E. / Bouwman, Jildau / Brookes, Antony J. / Clark, Tim / Crosas, Mercè / Dillo, Ingrid / Dumon, Oliver / Edmunds, Scott / Evelo, Chris T. / Finkers, Richard / Gonzalez-Beltran, Alejandra / Gray, Alasdair J.G. / Groth, Paul / Goble, Carole / Grethe, Jeffrey S. / Heringa, Jaap / A.C't Hoen, Peter / Hooft, Rob / Kuhn, Tobias / Kok, Ruben / Kok, Joost / Lusher, Scott J. / Martone, Maryann E. / Mons, Albert / Packer, Abel L. / Persson, Bengt / Rocca-Serra, Philippe / Roos, Marco / van Schaik, Rene / Sansone, Susanna-Assunta / Schultes, Erik / Sengstag, Thierry / Slater, Ted / Strawn, George / Swertz, Morris A. / Thompson, Mark / van der Lei, Johan / van Mulligen, Erik / Velterop, Jan / Waagmeester, Andrea / Wittenburg, Peter / Wolstencroft, Katherine / Zhao, Jun / Mons Barend (2016): „The FAIR Guiding Principles for scientific data management and stewardship“ in: *Scientific Data* 3, Article number: 160018. DOI: <https://doi.org/10.1038/sdata.2016.18>.